

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 24, 2002, 16:24:42 ; Search time 1810.21 Seconds  
(without alignments)  
16077.036 Million cell updates/sec

Title: US-09-708-724A-3\_COPY\_1\_1000

Perfect score: 1000

Sequence: 1 agccagactaggagtgcgc.....cacacatagagtcacagagga 1000

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*\*

- 1: gb\_ba:\*\*
- 2: gb\_htg:\*\*
- 3: gb\_in:\*\*
- 4: gb\_ov:\*\*
- 5: gb\_ov:\*\*
- 6: gb\_pat:\*\*
- 7: gb\_ph:\*\*
- 8: gb\_pl:\*\*
- 9: gb\_pr:\*\*
- 10: gb\_ro:\*\*
- 11: gb\_sts:\*\*
- 12: gb\_sy:\*\*
- 13: gb\_un:\*\*
- 14: gb\_vl:\*\*
- 15: em\_ba:\*\*
- 16: em\_fun:\*\*
- 17: em\_hum:\*\*
- 18: em\_in:\*\*
- 19: em\_mu:\*\*
- 20: em\_or:\*\*
- 21: em\_or:\*\*
- 22: em\_ov:\*\*
- 23: em\_pat:\*\*
- 24: em\_ph:\*\*
- 25: em\_pl:\*\*
- 26: em\_ro:\*\*
- 27: em\_sts:\*\*
- 28: em\_un:\*\*
- 29: em\_vl:\*\*
- 30: em\_htg\_hum:\*\*
- 31: em\_htg\_inv:\*\*
- 32: em\_htg\_other:\*\*
- 33: em\_htg\_mus:\*\*
- 34: em\_htg\_pln:\*\*
- 35: em\_htg\_rod:\*\*
- 36: em\_htg\_mam:\*\*
- 37: em\_htg\_vrt:\*\*
- 38: em\_sy:\*\*
- 39: em\_htgo\_hum:\*\*
- 40: em\_htgo\_mus:\*\*
- 41: em\_htgo\_other:\*\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1000	100.0	151049	9	AC018558	Homo sapi
2	1000	100.0	191768	9	AC092357	Homo sapi
3	306.4	30.6	167133	9	AC026186	Homo sapi
c 4	306.4	30.6	201061	9	AC007353	Homo sapi
c 5	279.8	28.0	144577	9	AC023824	Homo sapi
c 6	269.4	26.9	70048	9	HSU91320	Human chrom
c 7	269.4	26.9	129290	2	AC020761	Homo sapi
c 8	269.4	26.9	167553	2	AC114938	Homo sapi
c 9	269.4	26.9	185994	2	AC002042	Homo sapi
c 10	269.4	26.9	214025	9	AC007882	Homo sapi
c 11	269.4	26.9	227856	2	AC007908	Homo sapi
c 12	232.2	23.2	129290	2	AC020761	Homo sapi
c 13	121	12.1	57476	2	AC107630	Homo sapi
c 14	89	8.9	176547	9	AC092447	Homo sapi
c 15	87.4	8.7	194142	9	AC092634	Homo sapi
c 16	85.2	8.5	44013	2	AC087356	Homo sapi
c 17	79.8	8.0	104660	9	AP000347	AP000347 Pan trogl
c 18	79.8	8.0	208825	2	AC119407	Homo sapi
c 19	78	7.8	175952	9	AC069285	Homo sapi
c 20	66.4	6.6	57476	2	AC107630	Homo sapi
c 21	52.2	5.2	125020	9	AF429315	AF429315 Homo sapi
c 22	47.4	4.7	7218	6	I66494	Sequence 14 I66494
c 23	42.8	4.3	199654	2	AC010898	Homo sapi
c 24	42	4.2	10732	6	E32986	Gene encodi E32986
c 25	41.6	4.2	125020	9	AF429315	Homo sapi
c 26	41.4	4.1	157197	2	AC099042	Oryza sat AC099042
c 27	40.4	4.0	227856	2	AC007908	Homo sapi
c 28	40.2	4.0	94682	2	AC128917	Rattus no AC128917
c 29	40	4.0	165432	9	AL512590	Human DNA AL512590
c 30	39.8	4.0	65581	9	HSJ38L1	Human DNA AL109740
c 31	39.8	4.0	210344	9	AC007100	Homo sapi
c 32	39.4	3.9	69205	2	AC120034	Homo sapi
c 33	39.4	3.9	160584	2	AC022248	Homo sapi
c 34	39.2	3.9	230884	2	AC079523	Mus muscu AC079523
c 35	38.8	3.9	78986	9	AL354693	Human DNA AL354693
c 36	38.8	3.9	115666	2	AC105744	Oryza sat AC105744
c 37	38.8	3.9	163282	2	AC016438	Homo sapi
c 38	38.8	3.9	167553	2	AC114938	Homo sapi
c 39	38.8	3.9	168435	2	AL390764	Homo sapi
c 40	38.8	3.9	185304	2	AC106269	Rattus no AC106269
c 41	38.8	3.9	191014	9	AC113268	Papio cyn AC113268
c 42	38.6	3.9	146383	2	AC116367	Oryza sat AC116367
c 43	38.6	3.9	170261	2	AC008542	Homo sapi
c 44	38.6	3.9	173391	9	AC090063	Homo sapi
c 45	38.6	3.9	189076	2	AC023788	Homo sapi

ALIGNMENTS

RESULT 1  
AC018558

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AC018558 Homo sapiens chromosome 16 clone RP11-80F22, complete sequence.  
151049 bp DNA linear PRI 02-NOV-2001

AC018558

AC018558.5 GI:16596530

HTG.

Homo sapiens.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 151049)

DOE Joint Genome Institute.

Sequencing of Human Chromosome 16

Unpublished



Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

FEATURES

Location/Qualifiers  
1..191768  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="RP11-332P24"

BASE COUNT 57149 a 39592 c 40150 g 54877 t  
ORIGIN

Query Match 100.0%; Score 1000; DB 9; Length 191768;  
Best Local Similarity 100.0%; Pred. No. 1.3e-291;  
Matches 1000; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCCAGACTAGGAGTGACCCAGAGAGGGAAGGATGGTGAGGCACAGGCTGCACTCTA 60  
|||||  
Db 19966 AGCCAGACTAGGAGTGACCCAGAGAGGGAAGGATGGTGAGGCACAGGCTGCACTCTA 20025  
|||||

Qy 61 CTGGTGCCCCAGACCCAGACATGCATGCCAGGCTGCAGTCCAAAGGATACCTCGGTGCGGG 120  
Db 20026 CTGGTGCCCCAGACCCAGACATGCATGCCAGGCTGCAGTCCAAAGGATACCTCGGTGCGGG 20085  
|||||

Qy 121 TCCCTGTCCCCATAGCATCTAGATCAGCTGCTGAGGCTGGAGTCTTCCATTCCTTG 180  
Db 20086 TCCCTGTCCCCATAGCATCTAGATCAGCTGCTGAGGCTGGAGTCTTCCATTCCTTG 20145  
|||||

Qy 181 AGCATCAGGGTGTATCATCTTCCAAAGGTTTTCAGACATCCCTGTGACCCCTGGCA 240  
Db 20146 AGCATCAGGGTGTATCATCTTCCAAAGGTTTTCAGACATCCCTGTGACCCCTGGCA 20205  
|||||

Qy 241 GGGGCGGTTATCATGCGGATCGGTTCATGCGCTTCCTCCAAAGCAGACCCAGCAATCC 300  
Db 20206 GGGGCGGTTATCATGCGGATCGGTTCATGCGCTTCCTCCAAAGCAGACCCAGCAATCC 20265  
|||||

Qy 301 CCATGCCCAACCAATGCATCAATGTTTGTGGGCGCTTCTTGGGAAGCTCACCTTCTC 360  
Db 20266 CCATGCCCAACCAATGCATCAATGTTTGTGGGCGCTTCTTGGGAAGCTCACCTTCTC 20325  
|||||

Qy 361 CTCCTGTTTGGCCCTCCATCTTCCCAACCAAGTACTTCTGCGCATCCCTGTGTCACCA 420  
Db 20326 CTCCTGTTTGGCCCTCCATCTTCCCAACCAAGTACTTCTGCGCATCCCTGTGTCACCA 20385  
|||||

Qy 421 AATGGGAAATCGGTCTCGAGACTCAGAAACCACTGTGAGGCGCTCGAGTCTTCCCT 480  
Db 20386 AATGGGAAATCGGTCTCGAGACTCAGAAACCACTGTGAGGCGCTCGAGTCTTCCCT 20445  
|||||

Qy 481 GTCCTGGCTACAGGGCATGAATCAGAGAGAAAGTATCTTCCACCTCTCGAAGGCTG 540  
Db 20446 GTCCTGGCTACAGGGCATGAATCAGAGAGAAAGTATCTTCCACCTCTCGAAGGCTG 20505  
|||||

Qy 541 CCAGGCTCAGGCTTGGGCACACTGAGGCTGACAGGGGCTTCTGAAGGCCAGAGAGATG 600  
Db 20506 CCAGGCTCAGGCTTGGGCACACTGAGGCTGACAGGGGCTTCTGAAGGCCAGAGAGATG 20565  
|||||

Qy 601 GCCCGGACATAGGCTGAAGCAACCTGTCTGAGCCAAAGATCTGTTGTGCTCTCTCTGA 660  
Db 20566 GCCCGGACATAGGCTGAAGCAACCTGTCTGAGCCAAAGATCTGTTGTGCTCTCTCTCTGA 20625  
|||||

Qy 661 ATCTTAGTGGCTTCTTAAAGCGGGTGTGATCAGCCATGGGTATCAGACACTGGAGTCT 720  
|||||

Db 20626 ATCTTAGTGGCTTCTTAAAGCGGGTGTGATCAGCCATGGGTATCAGACACTGGAGTCT 20685  
Qy 721 CAGTAGCTCTAGGTGGGACACGGGCACAAATTTCACTTGCAGACCAGCTGCACGAGTGG 780  
|||||  
Db 20686 CAGTAGCTCTAGGTGGGACACGGGCACAAATTTCACTTGCAGACCAGCTGCACGAGTGG 20745  
|||||

Qy 781 ATAAAGAGAGAGTCTGTGTGGGAATCCTCTTGTGGTGCATCATCAGGAGGTGAAGTCTT 840  
Db 20746 ATAAAGAGAGAGTCTGTGTGGGAATCCTCTTGTGGTGCATCATCAGGAGGTGAAGTCTT 20805  
|||||

Qy 841 TGTCTATAGCCCTCATATCCAGCTTGTGTATACCAATTCAGTGAAGCTGGACAACAGCTGG 900  
Db 20806 TGTCTATAGCCCTCATATCCAGCTTGTGTATACCAATTCAGTGAAGCTGGACAACAGCTGG 20865  
|||||

Qy 901 CACTGCTCAACAGCGCTTACCAAGACATCATGTTTTTTTTTTTTTTTCCACCAACCTG 960  
Db 20866 CACTGCTCAACAGCGCTTACCAAGACATCATGTTTTTTTTTTTTTTTCCACCAACCTG 20925  
|||||

Qy 961 GACCTGAATGGGATGTGGACACACATAGTCCAGAGGA 1000  
Db 20926 GACCTGAATGGGATGTGGACACACATAGTCCAGAGGA 20965  
|||||

RESULT 3  
AC026186 167133 bp DNA linear HTG 29-MAY-2000  
LOCUS Homo sapiens chromosome 3 clone RP11-303p8 map 3p, WORKING DRAFT  
DEFINITION AC026186  
SEQUENCE, 44 unordered pieces.  
ACCESSION AC026186  
VERSION AC026186.2 GI:8101183  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 167133)  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H.,  
Dong, W., Fan, H., Feng, X., Guan, Q., Gu, X., Guo, D., He, L., Hu, S.,  
Huang, F., Jin, Y., Kang, N., Li, C., Li, G., Li, J., Li, L.,  
Li, S., Li, T., Liu, Y., Liu, B., Liu, Y., Li, W., Li, Y.,  
Luo, J., Niu, Y., Qi, Q., Qi, X., Song, S., Sun, M., Sun, Y.,  
Tao, R., Wang, H., Wang, J., Wang, L., Wang, L., Wang, R.,  
Wang, X., Wang, X., Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y.,  
Yan, C., Yang, X., Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, H.,  
Zhang, L., Zhang, M., Zhang, X., Zhang, X., Zhang, Y., Zhang, Y.,  
Zhang, Z., Zhu, B., Yu, J. and Yang, H.

Chromosome 3p genomic sequence  
Unpublished  
JOURNAL 2 (bases 1 to 167133)  
REFERENCE  
AUTHORS Chen, J., Hu, S., Dong, W., Wang, J., Zhang, Y., Zhang, H., Liu, B.,  
Bao, W., Sun, Y., Wu, Q., Wang, H., Yang, X., Cheng, C., Wang, Y., Niu, Y.,  
Qi, X., Li, T., Zhang, H., Liu, N., Wu, D., Yu, B., Fan, H., Liu, Y.,  
Li, G., Li, C., Bao, Q., Bao, J., Wang, X., Song, L., Zhang, L., Guo, D.,  
Huang, F., Zhang, G., Li, J., Bian, X., Zhang, W., Li, L., Feng, X., Yu, J.  
and Yang, H.

Direct Submission  
Submitted (21-MAR-2000) Human Genomic Center, Institute of  
Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing  
100101, P.R.China  
On May 29, 2000 this sequence version replaced gi:7272003.  
Center:Beijing Center  
Website: http://hgsc.igtp.ac.cn  
http://www.genomics.org.cn  
Contact: hgc@igtp.ac.cn

Center code: Beijing  
Project Information  
Center project name: 1% project  
Center clone name: RP11-303p8  
Summary Statistics  
Sequencing vector: pUC18; 100% of reads  
Chemistry: Dye-terminator; Et 55% of reads  
Chemistry: Dye-terminator Big Dye; 45% of reads

Assembly program: Phrap: version 0.990329  
Consensus quality: 144681 bases at least Q40  
Consensus quality: 152516 bases at least Q30  
Consensus quality: 159163 bases at least Q20  
Insert size: 121837; sum-of-contigs  
Quality coverage: 4.43x in Q20 bases; sum-of-contigs  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 44 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 2054: contig of 2054 bp in length  
\* 2055 2154: gap of unknown length  
\* 2155 3414: contig of 1260 bp in length  
\* 3415 3514: gap of unknown length  
\* 3515 5281: contig of 1767 bp in length  
\* 5282 5381: gap of unknown length  
\* 5382 6678: contig of 1297 bp in length  
\* 6679 6778: gap of unknown length  
\* 6779 8533: contig of 1755 bp in length  
\* 8534 8634: gap of unknown length  
\* 8634 9989: contig of 1356 bp in length  
\* 9990 10089: gap of unknown length  
\* 10090 11298: contig of 1209 bp in length  
\* 11299 11398: gap of unknown length  
\* 11399 12931: contig of 1533 bp in length  
\* 12932 13031: gap of unknown length  
\* 13032 14756: contig of 1725 bp in length  
\* 14757 14856: gap of unknown length  
\* 14857 16838: contig of 1982 bp in length  
\* 16839 16938: gap of unknown length  
\* 16939 18630: contig of 1692 bp in length  
\* 18631 18730: gap of unknown length  
\* 18731 20330: contig of 1600 bp in length  
\* 20331 20430: gap of unknown length  
\* 20431 21642: contig of 1212 bp in length  
\* 21643 21742: gap of unknown length  
\* 21743 23363: contig of 1621 bp in length  
\* 23364 25148: gap of unknown length  
\* 25149 25248: gap of unknown length  
\* 25249 27496: contig of 2248 bp in length  
\* 27497 27596: gap of unknown length  
\* 27597 30050: contig of 2454 bp in length  
\* 30051 30150: gap of unknown length  
\* 30151 32351: contig of 2201 bp in length  
\* 32352 32451: gap of unknown length  
\* 32452 34424: contig of 1973 bp in length  
\* 34425 34524: gap of unknown length  
\* 34525 36830: contig of 2306 bp in length  
\* 36831 36930: gap of unknown length  
\* 36931 39686: contig of 2756 bp in length  
\* 39687 39786: gap of unknown length  
\* 39787 42512: contig of 2726 bp in length  
\* 42513 42612: gap of unknown length  
\* 42613 44556: contig of 1944 bp in length  
\* 44557 44656: gap of unknown length  
\* 44657 48395: contig of 3739 bp in length  
\* 48396 48495: gap of unknown length  
\* 48496 50935: contig of 2440 bp in length  
\* 50936 51035: gap of unknown length  
\* 51036 54435: contig of 3400 bp in length  
\* 54436 54535: gap of unknown length  
\* 54536 59749: contig of 5214 bp in length  
\* 59750 59849: gap of unknown length  
\* 59850 63021: contig of 3172 bp in length  
\* 63022 63121: gap of unknown length  
\* 63122 66858: contig of 3737 bp in length  
\* 66859 66958: gap of unknown length

\* 66959 71348: contig of 4390 bp in length  
\* 71349 71448: gap of unknown length  
\* 71449 75783: contig of 4335 bp in length  
\* 75784 75883: gap of unknown length  
\* 75884 79500: contig of 3617 bp in length  
\* 79501 83516: gap of unknown length  
\* 83517 83616: contig of 3916 bp in length  
\* 83617 88942: gap of unknown length  
\* 88943 89042: contig of 5326 bp in length  
\* 89043 95244: contig of 6202 bp in length  
\* 95245 95344: gap of unknown length  
\* 95345 99266: contig of 3922 bp in length  
\* 99267 99366: gap of unknown length  
\* 99367 104303: contig of 4937 bp in length  
\* 104304 104403: gap of unknown length  
\* 104404 110029: contig of 5626 bp in length  
\* 110030 110129: gap of unknown length  
\* 110130 117586: contig of 7457 bp in length  
\* 117587 117686: gap of unknown length  
\* 117687 125138: contig of 7452 bp in length  
\* 125139 125238: gap of unknown length  
\* 125239 132290: contig of 7052 bp in length  
\* 132291 132390: gap of unknown length  
\* 132391 142718: contig of 10328 bp in length  
\* 142719 142818: gap of unknown length  
\* 142819 154592: contig of 11774 bp in length  
\* 154593 154692: gap of unknown length  
\* 154693 167133: contig of 12441 bp in length.

## FEATURES

Location/Qualifiers  
1..167133  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="3"  
/map="3p"  
/clone="RP11-303P8"  
1..2054  
/note="assembly\_name:Contig17"  
2155..3414  
/note="assembly\_name:Contig18"  
3515..5281  
/note="assembly\_name:Contig19"  
5282..6678  
/note="assembly\_name:Contig21"  
6679..8533  
/note="assembly\_name:Contig22"  
8534..9989  
/note="assembly\_name:Contig25"  
10090..11298  
/note="assembly\_name:Contig27"  
11299..12931  
/note="assembly\_name:Contig28"  
12932..14756  
/note="assembly\_name:Contig29"  
14757..16838  
/note="assembly\_name:Contig30"  
16839..18630  
/note="assembly\_name:Contig31"  
18631..20330  
/note="assembly\_name:Contig32"  
20331..21642  
/note="assembly\_name:Contig33"  
21643..23363  
/note="assembly\_name:Contig34"  
23364..25148  
/note="assembly\_name:Contig35"  
25149..27496  
/note="assembly\_name:Contig36"  
27497..30050  
/note="assembly\_name:Contig37"  
30051..32351  
/note="assembly\_name:Contig38"  
32352..34424  
/note="assembly\_name:Contig39"

/note="assembly\_name:Contig39"  
34525..36830  
/note="assembly\_name:Contig40"  
36931..39686  
/note="assembly\_name:Contig41"  
39787..42512

misc\_feature

misc\_feature

misc\_feature

Query Match 30.6%; Score 306.4; DB 2; Length 167133;  
Best Local Similarity 76.6%; Pred. No. 2.6e-81;  
Matches 418; Conservative 0; Mismatches 116; Indels 12; Gaps 3;

QY 1 AGCCAGACTAGGAGTGAGCCAGAGAGGGAAGGATGTGGAGGCACAGGCTGCACTCT- 59  
|||||  
Db 149165 AGCCACGGAGAGTGAGCCAGAGAGAACACAGGCGTGGAGGTGCGGCTGCTCTC 149224  
  
QY 60 ACTGTGTCCCGACAGCCAGACTGCATGCCAGGCTGCAGTCCAAAGGATACCTCGGTGGG 119  
|||||  
Db 149225 ACTGTGTCCCGACAGACTAGGCTGTCTCCAGGCTGCCGTACAGGGGTACTCGGTGG 149284  
  
QY 120 GTCCCTGTCCCGATAGCATCTTAGATCAGCTGCCTGAGGCTGGAGCTTCTTCCATTCTT 179  
|||||  
Db 149285 GTCCAGGCGCCCTCAGTGGCCTGAGTGATTGGCTGCAGCTGCAGCTTCTTCCAGTTACT 149344  
  
QY 180 GAGCATCAGGGGTGTGTATCAATTTCCAAAGGTTTTCAGACAAATCCCTGTGACCCCTGGC 239  
|||||  
Db 149345 GAGCAACAGGGGCTGTATCATTTCCAAAGGATTTTCAGACAAATCCCGGTGACTCTTGGC 149404  
  
QY 240 AGGGGCGGTATCATATGCGGATC-----GGTCCATGGCTTGCCTCCAAAGCAGCACC 291  
|||||  
Db 149405 AGAATGTGGTATTATTTGGCGATCTCCAGAGGCGACAGGGCTTGCCTCCAAAGTACACC 149464  
  
QY 292 CAGCAATCCCGATGCCACCAATGCACTAAATGTTTGTGGGCTCTTTCTGGAAGCT 351  
|||||  
Db 149465 TAGCAATATCCCGTCCAGCCAGGCACTAAGCATTTGTGTAGGCTCTTCTGGAAGCT 149524  
  
QY 352 CACCTTCTCTCTCTGTTTGGCCCTCCATCTTCCCAAGCAGTACTTCTGGCCATCTCTCT 411  
|||||  
Db 149525 CACCTT---CTTCTCTCACCTATGTCTTCCCAAGGCTTCTTGGCCATCTCTCTCT 149581  
  
QY 412 TGTCAACCAATGGGAAACATGGGCTTGGAGCTCAGAACCACTGTGCGAGGCTTCGAG 471  
|||||  
Db 149582 TGTCAACCAATGGGAAACATGGGCTTGGAGCTCAGAACCACTGTGCGAGGCTTCGAG 149641  
  
QY 472 TCTTCCCTCTCTCTGCTTAACAGGCGATCGAATCAAGAGAGAAAGTCACTTCTCCACCTCC 531  
|||||  
Db 149642 TCTTCCCGAGCCAGCCACAGGGTGTGAATTCAGAGAAACATCATCTTCTCTCTCT 149701  
  
QY 532 TGAAGG 537  
|||||  
Db 149702 TGAAGG 149707

RESULT 4  
AC007353/c 201061 bp DNA linear PRI 01-JUN-2001  
LOCUS Homo sapiens chromosome 16 clone RP11-488I20, complete sequence.  
DEFINITION AC007353  
ACCESSION AC007353  
VERSION AC007353.4 GI:14277247  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens

Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Eukaryota; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 201061)  
DOE Joint Genome Institute.  
TITLE Sequencing of Human Chromosome 16  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 201061)  
AUTHORS Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,  
Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,  
Bryant,J., Tesmer,J., Meinkne,L., Longmire,J., White,S., Tatum,O.,  
Campbell,C., Fawcett,J., Maltbie,M., Bussod,M., Sutherland,R.,  
McMurry,K., Han,C. and Deaven,L.

TITLE  
JOURNAL

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

COMMENT

Direct Submission  
Submitted (22-APR-1999) Center for Human Genome Studies, DOE Joint  
Genome Institute, Los Alamos National Laboratory, MS M888, Los  
Alamos, NM 87545, USA  
3 (bases 1 to 201061)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (01-JUN-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On Jun 1, 2001 this sequence version replaced gi:9795555.

-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----

Sequence Quality Assessment:

This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the  
GenBank flat file format but are available as part  
of this entry's ASN.1 file.

FEATURES  
source

Location/Qualifiers  
1..201061  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="RP11-488I20"

BASE COUNT 57062 a 41994 c 41012 g 60993 t  
ORIGIN

Query Match 30.6%; Score 306.4; DB 9; Length 201061;  
Best Local Similarity 76.6%; Pred. No. 2.6e-81;  
Matches 418; Conservative 0; Mismatches 116; Indels 12; Gaps 3;

QY 1 AGCCAGACTAGGAGTGAGCCAGAGGGAAGGATGTGGAGGCACAGGCTGCACTCT- 59  
|||||  
Db 86787 AGCCACGGAGAGTGAGCCAGAGAGAACACAGGCGTGGAGGTGCGGCTGCTCTCTC 86728  
  
QY 60 ACTGTGTCCCGACAGCCAGACTGCATGCCAGGCTGCAGTCCAAAGGATACCTCGGTGGG 119  
|||||  
Db 86727 ACTGTGTCCCGACAGACTAGGCTGTCTCCAGGCTGCCGTACAGGGGTACTCGGTGG 86668  
  
QY 120 GTCCCTGTCCCGATAGCATCTTAGATCAGCTGCCTGAGGCTGGAGCTTCTTCCATTCTT 179  
|||||  
Db 86667 GTCCAGGCGCCCTCAGTGGCCTGAGTGATTGGCTGCAGCTGCAGCTTCTTCCAGTTACT 86608  
  
QY 180 GAGCATCAGGGGTGTGTATCATTTCCAAAGGTTTTCAGACAAATCCCTGTGACCCCTGGC 239  
|||||  
Db 86607 GAGCAACAGGGGCTGTATCATTTCCAAAGGATTTTCAGACAAATCCCGGTGACTCTTGGC 86548  
  
QY 240 AGGGGCGGTATCATATGCGGATC-----GGTCCATGGCTTGCCTCCAAAGCAGCACC 291  
|||||  
Db 86547 AGAATGTGGTATTATTTGGCGATCTCCAGAGGCGACAGGCTTGCCTCCAAAGTACACC 86488  
  
QY 292 CAGCAATCCCGATGCCACCAATGCACTAAATGTTTGTGGGCTCTTCTTCTGGAAGCT 351  
|||||  
Db 86487 TAGCAATATCCCGTCCAGCCAGGCACTAAGCATTTGTGTAGGCTCTTCTTCTGGAAGCT 86428  
  
QY 352 CACCTTCTCTCTCTGTTTGGCCCTCCATCTTCCCAAGCAGTACTTCTGGCCATCTCTCT 411  
|||||  
Db 86427 CACCTT---CTTCTCTCACCTATGTCTTCCCAAGGCTTCTTCTGGCCATCTCTCTCT 86371  
  
QY 412 TGTCAACCAATGGGAAACATGGGCTCGGAGCTCAGAACCACTGTGCGAGGCTTCGAG 471  
|||||  
Db 86370 TGTCAACCAATGGGAAACATGGGCTCGGAGCTCAGAACCACTGTGCGAGGCTTCGAG 86311  
  
QY 472 TCTTCCCTCTCTCTGCTTAACAGGCGATCGAATCAGAGAGAAAGTCACTTCTTCCACCTCC 531  
|||||  
Db 86310 TCTTCCCGAGCCAGCCACAGGGTGTGAATTCAGAGAGAAACATCATCTTCTTCTCTCT 86251





```
repeat_region complement(30864..31123)
repeat_region /rpt_family="AluJo"
repeat_region 32213..32238
repeat_region /rpt_family="GC-rich"
repeat_region 32353..32518
repeat_region /rpt_family="(CGG)n"
repeat_region complement(32639..32781)
repeat_region /rpt_family="(TA)n"
repeat_region complement(32728..32877)
repeat_region /rpt_family="(TAA)n"
repeat_region 33451..33494
repeat_region /rpt_family="(CA)n"
repeat_region 34354..34767
repeat_region /rpt_family="L1"
repeat_region 34786..34917
repeat_region /rpt_family="FLAM_A"
repeat_region 35026..35342
repeat_region /rpt_family="L1"
repeat_region complement(35555..36427)
repeat_region /rpt_family="L1PA16"
repeat_region complement(36279..36847)
```

```
Query Match 26.9%; Score 269.4; DB 9; Length 70048;
Best Local Similarity 76.5%; Pred. No. 4.3e-70;
Matches 368; Conservative 0; Mismatches 86; Indels 27; Gaps 2;
```

```
QY 525 CACCTCTGAAGCGTCCAGCGCTGCGAGGCTGCGACACTGAGCGCTGACAGGGGCGCTTCG 584
Db 51193 CCCACGGAGAGACCGCTAGGCTGAGGCTTGTCACGTGAGCGCTGGCAGGGCCCTTCGG 51252
QY 585 AAGGCCAGAGGAGTGCCCGGGACATAGGCTGAAGCAACCTGCTGTGAGCCAAAGATCT 644
Db 51253 AAGGCAAGGGAAGATGGCCAGGACACACAAAGGCTGAAGCAACCCATCTGAGCCAAAGATCT 51312
QY 645 GTTGTGTCTCTCTGATCTTGTAGTGGCTCTTAAGGCGGGTGTGATCAGCCATGGGTAT 704
Db 51313 GTTGTAGTCTCTCTGATCTGATCTCAGAGCCTTGCCAAAGGAAGCATGATTACCCATGGGTAT 51372
QY 705 CAGAGACACTGAGTGCAGTACGTCTAGTGGG-----ACACGGGC 746
Db 51373 CAGAGACACTGAGGCTGCACCTGGCGGTGGCCAGGAGGCGCTGACTACTGCACAGT 51432
QY 747 ACAATTTCACTTGCAGACAGCTGCACGGAGTGGATAAGAGAGAGTCTGTGTGGGAAT 806
Db 51433 GTGATTTCACTGCGAGGCCAACTGCAGGGAGTGGATAAGAGAGAGCTCTGTGTGGGAAT 51492
QY 807 CTCCTTTGTGTGATCATCAGGAGGTGAGTCTTTGTATACGCTCATATCCAGCTTGTG 866
Db 51493 CTCCTTCGTGTGATCATCAGGAGGTGAGTCTTTGTATACGCTCATATCCAGCTTGTG 51552
QY 867 TGATACCAATTCCAGTGAAGCTGGAAAGCTGGCACTGCTCAACAGGCGCTTACCAAGAC 926
Db 51553 GATAGCAATTCAGTGAAGCTGGGACAGCTGGCACTGCTCAACAGGCGCTTCCCAAGAT 51612
QY 927 ATCATGTTTTTTTTTTTTTTTCCACCAACCTGACCTGAATGGGATGTGGACACACA 986
Db 51613 ACCAGG-----TTTCTTTCCACACAGCTGGGCGCTGGATGGGATGTGAACATCTA 51663
QY 987 T 987
Db 51664 T 51664
```

```
RESULT 7
AC020761/c AC020761 129290 bp DNA linear HTG 25-APR-2001
LOCUS Homo sapiens chromosome 16 clone RP11-151M15, WORKING DRAFT
DEFINITION SEQUENCE, 20 unordered pieces.
ACCESSION AC020761
VERSION AC020761.5 GI:9965531
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 129290)  
DOE Joint Genome Institute.  
Sequencing of Human Chromosome 16  
Unpublished  
2 (bases 1 to 129290)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (09-JAN-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On Sep 3, 2000 this sequence version replaced gi:9098615.

-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: <http://www.jgi.doe.gov>

Project Information  
Center Project Name: 0  
Center clone name: RPCI-11\_151M15

#### Summary Statistics

Consensus quality: 96142 bases at least Q40  
Consensus quality: 106591 bases at least Q30  
Consensus quality: 112022 bases at least Q20  
Estimated insert size: 161300; agarose-fp estimation  
Estimated insert size: 127390; sum-of-contigs estimation  
Quality coverage: 4.28 in Q20 bases; agarose-fp estimation  
Quality coverage: 5.42 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 20 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1	1043: contig of 1043 bp in length
1044	1143: gap of unknown length
1144	2482: contig of 1339 bp in length
2483	2582: gap of unknown length
2583	3593: contig of 1011 bp in length
3594	3693: gap of unknown length
3694	4803: contig of 1110 bp in length
4804	4903: gap of unknown length
4904	6120: contig of 1217 bp in length
6121	6220: gap of unknown length
6221	7568: contig of 1348 bp in length
7569	7668: gap of unknown length
7669	9284: contig of 1616 bp in length
9285	9384: gap of unknown length
9385	10398: contig of 1014 bp in length
10399	10498: gap of unknown length
10499	12201: contig of 1703 bp in length
12202	12301: gap of unknown length
12302	13971: contig of 1670 bp in length
13972	14071: gap of unknown length
14072	15687: contig of 1616 bp in length
15688	15787: gap of unknown length
15788	16900: contig of 1113 bp in length
16901	17000: gap of unknown length
17001	19643: contig of 2643 bp in length
19644	19743: gap of unknown length
19744	22031: contig of 2288 bp in length
22032	22131: gap of unknown length
22132	27299: contig of 5168 bp in length
27300	27399: gap of unknown length
27400	33138: contig of 5739 bp in length
33139	33238: gap of unknown length
33239	41651: contig of 8413 bp in length
41652	41751: gap of unknown length
41752	47053: contig of 5302 bp in length
47054	47153: gap of unknown length



```
* 47154 88215: contig of 41062 bp in length
* 88316 88315: gap of unknown length
* 88316 129290: contig of 40975 bp in length.
FEATURES
  Location/Qualifiers
    1..129290
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="16"
      /clone="RP11-151M15"
      /clone_lib="RPCI human BAC library 11"
BASE COUNT 34282 a 28372 c 28553 g 36024 t 2059 others
ORIGIN

Query Match      26.9%; Score 269.4; DB 2; Length 129290;
Best Local Similarity 76.5%; Pred. No. 4.3e-70;
Matches 368; Conservative 0; Mismatches 86; Indels 27; Gaps 2;

Qy 525 CACCTCTGAAGGCTGCAGCGCTGAGGCTTGGACACTGAGGTGACAGGGGCTTCTG 584
Db 69206 CCCACGGAGAGACCGCTAGGCTTGGTACAGTGAAGCTGAGGCTGGCAGGAGCTTCGG 69147

Qy 585 AAGCCAGAGAGATGGCCGGGACATAAGGCTGAAGCAACCTGCTGAGCCAAAGATCT 644
Db 69146 AAGCAAGGGAATGGCCAGGACACAAAGCTGAAGCAACCCATCTGAGCCAAAGATCT 69087

Qy 645 GTTTGTCTCTCCTGAATCTTTAGTGGCTTCTAAAGGGGGGTGTGATCAGCCATGGGTAT 704
Db 69086 GTTTAGGTCTCTTGAATCTCAGCAGCCTTCCCAAGGAAGCATGATTACCCATGGGTAT 69027

Qy 705 CAGACACACTGGATCCAGTACGTAGTGGG-----ACACGGGC 746
Db 69026 CAGACACACTGGAGCTGGCAGCTGGGGTGGCCAGAGGCTGACTACTGCACAGGT 68967

Qy 747 ACAATTTCACTTGACAGCACCTGCACGAGTGGATGAAGAGAGATTTCTGTGGGAAT 806
Db 68966 GTGATTTCACTGGCAGCCCACTCAGGAGTGGATGAAGAGAGCTCTGTGGGAAT 68907

Qy 807 CTCCTTTGGTGATCAGGAGGTGAAGCTTTTGTATAGCTTCATAGCTCATATCAGCTTGTG 866
Db 68906 CTCCTCGTGATCATCGAAGAGGTGAAGCTTCTTCATAGCTCAACCCCAATTTGTG 68847

Qy 867 TGATACCAATTCAGTGAAGCTGAACAAGCTGGCACTGTCTCAACAGCGCTTACCAAGAC 926
Db 68846 GGATAGCAATTCAGTGAAGCTGGGCAAGCTGGCACTGTCTCAACAGCGCTTCCCAAGAT 68787

Qy 927 ATCATGTTTTTTTTTTTTTTTCCACCAACCTGCACCTGAATGGGATGTGGACACACA 986
Db 68786 ACCAGG-----TTTCTTTCCACCACACAGCTGGGCTGGATTTGGGATGTGAACATCTA 68736

Qy 987 T 987
Db 68735 T 68735

RESULT 8
AC114938/c
LOCUS
DEFINITION Homo sapiens chromosome 5 clone CTD-2244D9, WORKING DRAFT SEQUENCE,
3 unordered pieces.
AC114938
VERSION AC114938.1 GI:19424399
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE
  Homo sapiens.
  ORGANISM
    Homo sapiens
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE
    1 (bases 1 to 167553)
      AUTHORS DOE Joint Genome Institute.
      TITLE Sequencing of Human Chromosome 5
      JOURNAL Unpublished
      REFERENCE
        2 (bases 1 to 167553)
          AUTHORS DOE Joint Genome Institute.
          TITLE Direct Submission

Submitted (14-MAR-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 721326
Center clone name: CITB-HI_2244D9
-----
Summary Statistics
Consensus quality: 163638 bases at least Q40
Consensus quality: 165018 bases at least Q30
Consensus quality: 166006 bases at least Q20
Estimated insert size: 150000; agarose-fp estimation
Estimated insert size: 167353; sum-of-contigs estimation
Quality coverage: 10.56 in Q20 bases; agarose-fp estimation
Quality coverage: 9.46 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 31475: contig of 31475 bp in length
* 31476 31575: gap of unknown length
* 31576 82160: contig of 50585 bp in length
* 82161 82260: gap of unknown length
* 82261 167553: contig of 85293 bp in length.
FEATURES
  Location/Qualifiers
    1..167553
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="5"
      /clone="CTD-2244D9"
      /clone_lib="Caltech human BAC library D"
BASE COUNT 48962 a 35722 c 32192 g 50416 t 261 others
ORIGIN

Query Match      26.9%; Score 269.4; DB 2; Length 167553;
Best Local Similarity 76.5%; Pred. No. 4.3e-70;
Matches 368; Conservative 0; Mismatches 86; Indels 27; Gaps 2;

Qy 525 CACCTCTGAAGGCTGCAGCGCTGAGGCTTGGACACTGAGGTGACAGGGGCTTCTG 584
Db 63442 CCCACGGAGAGACCGCTAGGCTTGGTACAGTGAAGCTGAGGCTGGCAGGAGCTTCGG 63383

Qy 585 AAGCCAGAGAGATGGCCGGGACATAAGGCTGAAGCAACCTGCTGAGCCAAAGATCT 644
Db 63382 AAGCAAGGGAATGGCCAGGACACAAAGCTGAAGCAACCCATCTGAGCCAAAGATCT 63323

Qy 645 GTTTGTCTCTCCTGAATCTTTAGTGGCTTCTAAAGGGGGGTGTGATCAGCCATGGGTAT 704
Db 63322 GTTTAGGTCTCTTGAATCTCAGCAGCCTTCCCAAGGAAGCATGATTACCCATGGGTAT 63263

Qy 705 CAGACACACTGGATCCAGTACGTAGTGGG-----ACACGGGC 746
Db 63262 CAGACACACTGGAGCTGGCAGCTGGGGTGGCCAGAGGCTGACTACTGCACAGGT 63203

Qy 747 ACAATTTCACTTGACAGCACCTGCACGAGTGGATGAAGAGAGATTTCTGTGGGAAT 806
Db 63202 GTGATTTCACTGGCAGGCCAACTCAGGAGTGGATGAAGAGAGAGCTCTGTGGGAAT 63143

Qy 807 CTCCTTTGGTGATCATCAGGAGGTGAAGCTTTTGTATAGCTTCATATCAGCTCATATCAGCTTGTG 866
Db 63142 CTCCTCGTGATCATCGAAGAGGTGAAGTCTTCTTCATAGCTCAACCCCAATTTGTG 63083

Qy 867 TGATACCAATTCAGTGAAGCTGGAACAAGCTGGCACTGTCTCAACAGCGCTTACCAAGAC 926
Db 63082 GGATAGCAATTCAGTGAAGCTGGGCAACAGCTGGCACTGTCTCAACAGCGCTTCCCAAGAT 63023
```





```
repeat_region 4920..4970 /rpt_family="MER2_type"
repeat_region 4974..5104 /rpt_family="MER2_type"
repeat_region 5116..5219 /rpt_family="MER2_type"
repeat_region 5334..5416 /rpt_family="MER2_type"
misc_feature 5602..5955 /note="similar to EST A1806989 (NID:g5393555) wf24h01.x1"
repeat_region 5824..5887 /rpt_family="MER103"
misc_feature 6368..6806 /note="similar to EST A1187426 (NID:g3738064) qf31d08.x1"
repeat_region 6368..6520 /rpt_family="Alu"
repeat_region 6886..7074 /rpt_family="MER1_type"
repeat_region 7184..7418 /rpt_family="MER1_type"

Query Match 26.9%; Score 269.4; DB 9; Length 214025;
Best Local Similarity 76.5%; Pred. No. 4.3e-70;
Matches 368; Conservative 0; Mismatches 86; Indels 27; Gaps 2;

QY 525 CACCTCTGAAGCGCCAGCGCTAGGCTTGGCAGCTGAGCGCTGACAGGGGCCCTTCGTG 584
Db 27450 CCCCACGGAGAGACCGCTAGGCTTGGTACAGCTGAGCGCTGGCAGGGCCCTTCGG 27509

QY 585 AAGGCCAGAGAGATGGCCCGGACATAGCTGAAGCAACCTGCTGAGCCCAAGATCT 644
Db 27510 AAGGCAGGGAAGATGGCCCAAGGACACAAAGCTGAAGCAACCCATCTGAGCCCAAGATCT 27569

QY 645 GTTGTGTCCTCCTGAATCTTAGTGCCCTTCTAAAGCGCGGTGTGATCAGCCATGGTAT 704
Db 27570 GTTGTGTCCTCCTGAATCTCAGCAGCCCTGCCAAGGAAGCATGATTACCCATGGTAT 27629

QY 705 CAGAGACACTGAGTCCAGTACTGCTAGGTGGG-----ACACGGGC 746
Db 27630 CAGAGACACTGAGGTGCGACGTGGCGGTGGCGGAGGAGCGCTGACTACTGTCACAGGT 27689

QY 747 ACAATTTCACTTGCACAGCAGCTGCACGGAGTGGATAAGAGAGAGTCTGTGTGGGAAT 806
Db 27690 GTGATTTCACTGCGAGGCCAATCTCAGGGAGTGGATAAGAGAGAGTCTGTGTGGGAAT 27749

QY 807 CTCCTTTGGTGGATCATCAGGAGGTGAAGTCTTTGTATAGCCCTCATATCCAGCTTGTG 866
Db 27750 CTCCTCGTGGTATCATCAGGAGGTGAAGTCTTCTCATAGCTCAACCCCAATTTGTG 27809

QY 867 TGATACCAATTCAGTGAAGCTGGACAAAGCTGGCAGCTGCTCAACAGGCGCTTACCAAGAC 926
Db 27810 GGATAGCAATTCAGTGAAGCTGGGACAAAGCTGGCAGCTGCTCAACAGGCGCTTCCCAAGAT 27869

QY 927 ATCATGTTTTTTTTTTTTTTTCCACCAACCTGGACCTGATGGGATGGGACACACA 986
Db 27870 ACCAGG-----TTTCTTCCACCAAGCTGGGCGCTGGATGGGATGGTGAACATCTA 27920

QY 987 T 987
Db 27921 T 27921

RESULT 11
AC007908/c
LOCUS AC007908 227856 bp DNA linear HTG 25-APR-2001
DEFINITION Homo sapiens chromosome 16 clone RP11-499D5, WORKING DRAFT
SEQUENCE, 30 unordered pieces.
ACCESSION AC007908
VERSION AC007908.3 GI:9094205
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

1 (bases 1 to 227856)  
DOE Joint Genome Institute.  
Sequencing of Human Chromosome 16  
Unpublished  
2 (bases 1 to 227856)  
Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,  
Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,  
Bryant,J., Tesmer,J., Meinkne,L., Longmire,J., White,S., Tatum,O.,  
Campbell,C., Fawcett,J., Malthie,M., Bussod,M., Sutherland,R.,  
McMurry,K., Han,C. and Deaven,L.  
Direct Submission  
Submitted (24-JUN-1999) Center for Human Genome Studies, DOE Joint  
Genome Institute, Los Alamos National Laboratory, MS M888, Los  
Alamos, NM 87545, USA  
On Jul 13, 2000 this sequence version replaced gi:7211873.  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----  
Project Information  
Center Project Name: 0  
Center clone name: RPCI-11\_499D5

\*\*\*\*\*  
Summary Statistics  
Consensus quality: 184734 bases at least Q40  
Consensus quality: 201642 bases at least Q30  
Consensus quality: 209314 bases at least Q20  
Estimated insert size: 218590; agarose-fp estimation  
Estimated insert size: 224956; sum-of-contigs estimation  
Quality coverage: 4.43 in Q20 bases; agarose-fp estimation  
Quality coverage: 4.3 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 30 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 1259: contig of 1259 bp in length  
\* 1260 1359: gap of unknown length  
\* 1360 2455: contig of 1096 bp in length  
\* 2456 2555: gap of unknown length  
\* 2556 3691: contig of 1136 bp in length  
\* 3692 3791: gap of unknown length  
\* 3792 4805: contig of 1014 bp in length  
\* 4806 4905: gap of unknown length  
\* 4906 5938: contig of 1033 bp in length  
\* 5939 6038: gap of unknown length  
\* 6039 7051: contig of 1013 bp in length  
\* 7052 7151: gap of unknown length  
\* 7152 8246: contig of 1095 bp in length  
\* 8247 8347: gap of unknown length  
\* 8347 9994: contig of 1647 bp in length  
\* 9994 10094: gap of unknown length  
\* 10094 11375: contig of 1282 bp in length  
\* 11376 11476: gap of unknown length  
\* 11476 12513: contig of 1038 bp in length  
\* 12514 12614: gap of unknown length  
\* 12614 13646: contig of 1032 bp in length  
\* 13646 13746: gap of unknown length  
\* 13746 15252: contig of 1507 bp in length  
\* 15253 15353: gap of unknown length  
\* 15353 16574: contig of 1221 bp in length  
\* 16574 16674: gap of unknown length  
\* 16674 18206: contig of 1532 bp in length  
\* 18206 18306: gap of unknown length  
\* 18306 19914: contig of 1608 bp in length  
\* 19914 20014: gap of unknown length  
\* 20014 22479: contig of 2466 bp in length  
\* 22480 22579: gap of unknown length  
\* 22580 22593: contig of 2714 bp in length

\* 25294 25393: gap of unknown length  
\* 25394 29118: contig of 3725 bp in length  
\* 29119 29218: gap of unknown length  
\* 32189: contig of 2971 bp in length  
\* 32190 32289: gap of unknown length  
\* 32290 32394: contig of 1005 bp in length  
\* 33395 33394: gap of unknown length  
\* 33395 34509: contig of 1115 bp in length  
\* 34510 34609: gap of unknown length  
\* 34610 34610: gap of unknown length  
\* 36148 36247: gap of unknown length  
\* 36248 42248: contig of 6001 bp in length  
\* 42249 42349: gap of unknown length  
\* 42349 52325: contig of 9977 bp in length  
\* 52326 52425: gap of unknown length  
\* 52426 60479: contig of 8054 bp in length  
\* 60480 60579: gap of unknown length  
\* 60580 68636: contig of 8057 bp in length  
\* 68637 68736: gap of unknown length  
\* 68737 83105: contig of 14369 bp in length  
\* 83106 83205: gap of unknown length  
\* 83206 116897: contig of 33692 bp in length  
\* 116898 118997: gap of unknown length  
\* 118998 165661: contig of 48564 bp in length  
\* 165662 165662: gap of unknown length  
\* 165662 227856: contig of 62195 bp in length.

FEATURES  
source  
1..227856  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="RP11-499D5"  
/clone.lib="RPCI human BAC library 11"  
BASE COUNT 68439 a 46078 c 44943 g 64859 t 3537 Others  
ORIGIN

Query Match 26.9%; Score 269.4; DB 2; Length 227856;  
Best Local Similarity 76.5%; Pred. No. 4.3e-70;  
Matches 368; Conservative 0; Mismatches 86; Indels 27; Gaps 2;  
QY 525 CACCTCCTGAAGCTGCCAGCGTCAAGCGCTTGGCACACTGAGGCTGAGCGGCGCTTCGTG 584  
Db 202358 CCCCACGAGAGACCGCTAGGCTGAGGCTTGGTACAGTGAAGCTGAGGCGGCGCTTCGG 202299  
QY 585 AAGGCCAGAGGATGGCCGGGACATAAGGCTGAAGCAACCTGCTCAGGCCAAGATCT 644  
Db 202298 AAGCCAAGGAGATGGCCAGGACACCAAGCTGAAGCAACCCATCTGAGCCAAAGATCT 202239  
QY 645 GTTTGTGCTCCTGAAATCTTTAGTGGCCTCTTAAAGCGGGTGTGATCAGCCATGGGTAT 704  
Db 202238 GTTTAGGTCTTCTGAATCTCAGCAGCCTTGCCAAAGGAAGCATGATTACCCATGGGTAT 202179  
QY 705 CAGACACACTGAGTCCAGTACTGCTAGTGGG-----ACACGGGC 746  
Db 202178 CAGACACACTGAGGCTGGCAGCTGGCGGTGGGCCAGGAGCGCTGACTACTGCACAGGT 202119  
QY 747 ACAATTTCACTTGCACGAGCTGCACGAGTGGGATAAAGAGAGAGTGTGTGTGGGAAT 806  
Db 202118 GTGATTTCACTGGCAGGCCAAGCTCAGGAGTGGGATAAAGAGAGAGCTGTGTGGGAAT 202059  
QY 807 CTCCTTTGGTGGATCATCAGGAGGTGAAGTCTTTTGTTCATAGCCTCATATCAGGCTTGTG 866  
Db 202058 CTCCTCGGTGGATCATCGAAGAGGTGAAGTCTTTCATAGCCTCAAAACCAATTTGTG 201999  
QY 867 TGATACCAATTCAGTGAAGCTGGNACAAAGCTGGCACTGCTCAACAGCGCTTACCAAGAC 926  
Db 201998 GGATAGCAATTCAGTGAAGCTGGNACAAAGCTGGCACTGCTCAACAGCGCTTCCCAAGAT 201939  
QY 927 ATCATGTTTTTTTTTTTTTTTCCACCACCAACCTGAGCCTGAATGGGGATGTGGACACACA 986  
Db 201938 ACCAGG-----TTTCTTTCCACCACAGCTGGGCGCTGGATTGGGATGTGAACATCTA 201888  
QY 987 T 987

Db 201887 T 201887

RESULT 12

AC020761

LOCUS

DEFINITION

AC020761

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AC020761 129290 bp DNA linear HTG 25-APR-2001  
Homo sapiens chromosome 16 clone RP11-151M15, WORKING DRAFT  
SEQUENCE, 20 unordered pieces.

AC020761.5 GI:9965531

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.

Homo sapiens.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 129290)

DOE Joint Genome Institute.

Sequencing of Human Chromosome 16

Unpublished

2 (bases 1 to 129290)

DOE Joint Genome Institute.

Direct Submission

Submitted (09-JAN-2000) Production Sequencing Facility, DOE Joint

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

On Sep 3, 2000 this sequence version replaced gi:9098615.

-----Genome Center

Center: Joint Genome Institute

Center Code: JGI

Web site: http://www.jgi.doe.gov

-----

Project Information

Center Project Name: 0

Center clone name: RPCI-11\_151M15

-----

Summary Statistics

Consensus quality: 96142 bases at least Q40

Consensus quality: 106591 bases at least Q30

Consensus quality: 112022 bases at least Q20

Estimated insert size: 161300; agarose-fp estimation

Estimated insert size: 127390; sum-of-contigs estimation

Quality coverage: 4.28 in Q20 bases; agarose-fp estimation

Quality coverage: 5.42 in Q20 bases; sum-of-contigs estimation.

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 20 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1

\* 1043: contig of 1043 bp in length

\* 1143: gap of unknown length

\* 1144 2482: contig of 1339 bp in length

\* 2483 2582: gap of unknown length

\* 2583 3593: contig of 1011 bp in length

\* 3594 3693: gap of unknown length

\* 3694 4803: contig of 1110 bp in length

\* 4804 4903: gap of unknown length

\* 4904 6120: contig of 1217 bp in length

\* 6121 7568: contig of 1348 bp in length

\* 7569 7668: gap of unknown length

\* 7669 9284: contig of 1616 bp in length

\* 9285 9385: gap of unknown length

\* 9385 10398: contig of 1014 bp in length

\* 10399 10498: gap of unknown length

\* 10499 12201: contig of 1703 bp in length

\* 12202 12301: gap of unknown length

\* 12302 13971: contig of 1670 bp in length

\* 13972 14071: gap of unknown length

\* 14072 15687: contig of 1616 bp in length

\* 15688 15787: gap of unknown length

\* 15788 16900: contig of 1113 bp in length  
\* 16901 17000: gap of unknown length  
\* 17001 19643: contig of 2643 bp in length  
\* 19644 19743: gap of unknown length  
\* 19744 22031: contig of 2288 bp in length  
\* 22032 22131: gap of unknown length  
\* 22132 27299: contig of 5168 bp in length  
\* 27300 27399: gap of unknown length  
\* 27400 33138: contig of 5739 bp in length  
\* 33139 33238: gap of unknown length  
\* 33239 41651: contig of 8413 bp in length  
\* 41652 41751: gap of unknown length  
\* 41752 47053: contig of 5302 bp in length  
\* 47054 47153: gap of unknown length  
\* 47154 88215: contig of 41062 bp in length  
\* 88216 88315: gap of unknown length  
\* 88316 129290: contig of 40975 bp in length.  
FEATURES  
    Location/Qualifiers  
        1..129290  
        /organism="Homo sapiens"  
        /db\_xref="taxon:9606"  
        /chromosome="16"  
        /clone="RP11-151M15"  
        /clone\_lib="RP11-151M15"  
BASE COUNT 34282 a 28372 c 28553 g 36024 t 2059 others  
ORIGIN  
Query Match 23.2%; Score 232.2; DB 2; Length 129290;  
Best Local Similarity 75.1%; Pred No. 8.1e-59;  
Matches 325; Conservative 0; Mismatches 86; Indels 20; Gaps 2;  
QY 525 CACCTCTGAAGGCTGCCAGCGTTCAGGCTTGCGACACTGAGGCTGACAGGGCCCTTCG 584  
Db 41178 CCCACGGAGAGACCACTAGGCTCAGGCTTGCTACAGTGAGGCTGGCAGGCGCCCTCGG 41237  
QY 585 AGGCCAGAGGATGGCCCGGACATAGGCTGAAGCAACCTGCTGAGCCAAAGATCT 644  
Db 41238 AAGGCAAGGGAAGATGGCCCAAGGACACAAAGGCTGAAGCAACCCATCTGAGCCAAAGATCT 41297  
QY 645 GTTGTGTCCTCTGAATCTTAGTGCCCTCTAAAGGCGGGTGTATGATGACCATGGGTAT 704  
Db 41298 GTTGTGTCCTCTGAATCTAGCAGCTGCGCAGGAAGGATGATTACCCATGGGTAT 41357  
QY 705 CAGAGACACTGGAGTCCAGTAGTCTGCTAGGTGGG-----ACACGGG 745  
Db 41358 CAGAGACACTGGAGGCTGCGAGCTGCGGGTGGCCAGGAGGCGCTGACTACTGCACAGG 41417  
QY 746 CACAATTTCACTTGCAGACAGCTGCACCGAGTGGATAAGAGAGAGTCTGTGTGGGAA 805  
Db 41418 TGTGATTTCACTTGGCAGGCAACTGCAGGAGTGGATAAGAGAGAGCTCTGTGTGGAA 41477  
QY 806 TCTCCTTTGTTGATCAGGAGGTGAAGTCTTTGTTCATAGCCCTCATATCCAGCTTCT 865  
Db 41478 TCTCCTCGTGTGATCATGAAGAAGTGAAGTCTTCTTCATAGCCCTTACCCCAATTTGT 41537  
QY 866 GTGATACCAATTCAGTGAAGCTGGAACAGCTGGCAGCTGCTCAACAGGCGCTTACCAAGA 925  
Db 41538 G-GATGCAATTCAGTGAACACTTGGACAAGCTGGCAGCTTGTTCACGAGCGCTTCCAAGA 41596  
QY 926 CATCATGTTTTTT 938  
Db 41597 TACCAGTTTCTTT 41609  
RESULT 13  
AC107630 AC107630 57476 bp DNA linear HTG 24-JAN-2002  
LOCUS Homo sapiens chromosome 17 clone CTD-2320G7 map 17, LOW-PASS  
DEFINITION SEQUENCE SAMPLING.  
ACCESSION AC107630  
VERSION AC107630.1 GI:18308327  
KEYWORDS HTGS\_PHASE0.  
SOURCE Homo sapiens.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
    1 (bases 1 to 57476)  
    Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Homo sapiens chromosome 17, clone CTD-2320G7  
    2 (bases 1 to 57476)  
    unpublished  
    Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhalter,B.,  
    Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,  
    Choepe,I., Colangelo,M., Collins,S., Collymore,A., Cook,A.,  
    Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,  
    Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
    Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
    Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
    Kamat,A., Karatas,A., Kellis,C., LaRoque,K., Lamazares,R.,  
    Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C.,  
    Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,  
    McEwan,P., McKernan,K., Meldrum,J., Meneus,L., Mihova,T.,  
    Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,  
    Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,  
    Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,  
    Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,  
    Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,  
    Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
    Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
    Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
    Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
    Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: tl2868  
Center clone name: 2320\_G\_7  
-----  
\* NOTE: This record contains 70 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.  
\* 1 699: contig of 699 bp in length  
\* 700 799: gap of 100 bp  
\* 800 1503: contig of 704 bp in length  
\* 1504 1603: gap of 100 bp  
\* 1604 2312: contig of 709 bp in length  
\* 2313 2412: gap of 100 bp  
\* 2413 3133: contig of 721 bp in length  
\* 3134 3233: gap of 100 bp  
\* 3234 3965: contig of 732 bp in length  
\* 3966 4065: gap of 100 bp  
\* 4066 4773: contig of 708 bp in length  
\* 4774 4873: gap of 100 bp  
\* 4874 5549: contig of 676 bp in length  
\* 5550 5649: gap of 100 bp  
\* 5650 6370: contig of 721 bp in length  
\* 6371 6470: gap of 100 bp  
\* 6471 7189: contig of 719 bp in length  
\*  
TITLE  
JOURNAL  
COMMENT

36997	37709:	contig of 713 bp	in length
37710	37809:	gap of 100 bp	in length
37810	38554:	contig of 745 bp	in length
38654	38654:	gap of 100 bp	in length
38655	39370:	contig of 716 bp	in length
39371	39470:	gap of 100 bp	in length
39471	40184:	contig of 714 bp	in length
40185	40284:	gap of 100 bp	in length
40285	41003:	contig of 719 bp	in length
41103	41103:	gap of 100 bp	in length
41104	41821:	contig of 718 bp	in length
41822	41921:	gap of 100 bp	in length
41922	42651:	contig of 730 bp	in length
42652	42751:	gap of 100 bp	in length
42752	43482:	contig of 731 bp	in length
43483	43582:	gap of 100 bp	in length
43583	44310:	contig of 728 bp	in length
44311	44410:	gap of 100 bp	in length
44411	45140:	contig of 730 bp	in length
45141	45240:	gap of 100 bp	in length
45241	45972:	contig of 732 bp	in length
45973	46072:	gap of 100 bp	in length
46073	46803:	contig of 731 bp	in length
46903	46903:	gap of 100 bp	in length
46904	47634:	contig of 731 bp	in length
47635	47734:	gap of 100 bp	in length
47735	48456:	contig of 722 bp	in length
48457	48556:	gap of 100 bp	in length
48557	49258:	contig of 702 bp	in length
49259	49358:	gap of 100 bp	in length
49359	50107:	contig of 749 bp	in length
50108	50207:	gap of 100 bp	in length
50208	50933:	contig of 726 bp	in length
50934	51033:	gap of 100 bp	in length
51034	51760:	contig of 727 bp	in length
51761	51860:	gap of 100 bp	in length
51861	52560:	contig of 700 bp	in length
52561	52660:	gap of 100 bp	in length
52661	53351:	contig of 691 bp	in length
53352	53451:	gap of 100 bp	in length
53452	54191:	contig of 740 bp	in length
54192	54291:	gap of 100 bp	in length
54292	55009:	contig of 718 bp	in length
55010	55109:	gap of 100 bp	in length
55110	55832:	contig of 713 bp	in length
55823	55922:	gap of 100 bp	in length
55923	56648:	contig of 726 bp	in length

Best Local Similarity 77.2%; Pred. No. 4.2e-25;

Matches	200;	Conservative	0;	Mismatches	50;	Indels	9;	Gaps	4;
---------	------	--------------	----	------------	-----	--------	----	------	----

Qy	290	CCAGCAATCCCATGCCCACCAATGACTAAATGTTTGTGGTGGGCTCTTTCTTGGGAAG	349
Db	20526	CCAGCATCCCGTGGCTACCGAGCAATTTGTGGTGGG--CTTTCTTGGGAAG	20583
Qy	350	CTCACTTCTCTCTGTTTGGCTTCCATCTTCCCAACACGACTACTCTTGGCCCATCTC	409
Db	20584	CTCGCTTCTCTGCTG--GCTCTTCTGTTTCCCAACCCGGTACTCTTGGTCATTCTC	20640
Qy	410	CTTG---TCACCACAATGGGAAATCGGTCTTGAGACTCAGAAACACACTGTGCGAGGC	466
Db	20641	CCTGTGATCATCACAAATGAGACAGCTGGTGGTGGAGACTCAGAAACTGCCATGCGAGCC	20700
Qy	467	TCGAGTCTTCCCTCTCTCTGGCTTAACAGGGCATGGAATCAGAGAAAGTCAATCTTCCA	526
Db	20701	TTGAGTCTTCCAGGCCACGCCACAGGGCAGAGAATCCAGGAAAGTCAT--TTTCA	20759
Qy	527	CCTCCTGAAGGCTGCCAGC	545
Db	20760	CCTCTGAAGGACCAGAGC	20778

```

RESULT 14
AC092447
LOCUS      176547 bp      DNA      linear      PRI 27-FEB-2002
DEFINITION Homo sapiens BAC clone RP11-760D2 from 7, complete sequence.
ACCESSION  AC092447
VERSION    AC092447.5  GI:18873962
KEYWORDS   HTG.
SOURCE     Homo sapiens.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 176547)
            Sulston, J.E. and Waterston, R.
            Toward a complete human genome sequence
            Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE    99063792
PUBMED     9847074
REFERENCE  2 (bases 1 to 176547)
            Armstrong, J. and Cotton, M.
            The sequence of Homo sapiens BAC clone RP11-760D2
            Unpublished (2001)
REFERENCE  3 (bases 1 to 176547)
            Waterston, R.H.
            Direct Submission
            Submitted (04-JUL-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE  4 (bases 1 to 176547)
            Waterston, R.H.
            Direct Submission
            Submitted (23-FEB-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE  5 (bases 1 to 176547)
            Waterston, R.
            Direct Submission
            Submitted (27-FEB-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
            On Feb 23, 2002 this sequence version replaced gi:16259201.
COMMENT    ----- Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc
            Contact: sapiens@watson.wustl.edu
            ----- Summary Statistics
            -----
            Center project name: H_NH0760D02
            -----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 7 Mapping
Project (Eric D. Green, Director), John D. McPherson in the
Department of Genetics (Washington University), and the Washington
University Genome Sequencing Center. For additional information
about the map position of this sequence, see
http://www.nhgri.nih.gov/DIR/CTB/CHR7 , send
mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male

```

```

donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E.,
Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at http://www.chori.org
VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP13-492C18; the clone sequenced
to the right is RP11-10F11, 2000 bp overlap. Actual start of this
clone is at base position 1 of RP11-760D2; actual end is at base
position 17615 of RP11-10F11.

The region from 20341 to 20385 is covered only by a PCR product
from clone DNA. An unresolved simple sequence repeats exists
between 84203 and 84724. Polymorphisms exist between AC024199 and
AC092447. Data from AC024199 and AC092423 was used to finish
AC092447.

FEATURES             Location/Qualifiers
     source            1..176547
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /map="7"
                        /chromosome="7"
                        /clone_lib="RPCI-11"
                        /clone="RP11-760D2"
     repeat_region     5..1938
                        /rpt_family="L1"
     repeat_region     2210..2483
                        /rpt_family="Alu"
     misc_feature       2799..3012
                        /note="match to EST BF894500 (NID: g12285959)"
     misc_feature       2799..2962
                        /note="match to EST BF895059 (NID: g12286518)"
     repeat_region     3092..3388
                        /rpt_family="L2"
     repeat_region     3419..3536
                        /rpt_family="L1"
     repeat_region     3537..3794
                        /rpt_family="Alu"
     repeat_region     3795..4263
                        /rpt_family="L1"
     repeat_region     4259..4542
                        /rpt_family="L1"
     repeat_region     4548..4748
                        /rpt_family="L1"
     repeat_region     4749..5046
                        /rpt_family="Alu"
     repeat_region     5047..5267
                        /rpt_family="L1"
     repeat_region     5268..5560
                        /rpt_family="Alu"
     repeat_region     5561..6091
                        /rpt_family="L1"
     repeat_region     6528..6825
                        /rpt_family="Alu"
     repeat_region     8380..8686
                        /rpt_family="Alu"
     repeat_region     8924..9225
                        /rpt_family="Alu"
     repeat_region     9838..10125
                        /rpt_family="Alu"
     repeat_region     10336..10600
                        /rpt_family="Alu"
     repeat_region     11975..12282
                        /rpt_family="Alu"
     repeat_region     12457..12765
                        /rpt_family="Alu"
     misc_feature       14411..14448
                        /note="match to EST AI334287 (NID: g4070846) qq25e08.x1"
     misc_feature       14411..14446
                        /note="similar to Homo sapiens EST BF979990"

```



Qy	951	ACCAAACTGACCTGAATGGGGATGTGGACA-CACATAGAGTCCAGAGGA	1000 
Db	49531	ACCATAGCTGAGCCTGGATGGATGTGGACACCTCAGACTCAGAGGA	49581
RESULT_15			
AC092634			
LOCUS	AC092634	194142 bp	DNA linear PRI 01-MAR-2002
DEFINITION	Homo sapiens BAC clone RP11-34016 from 7, complete sequence.		
ACCESSION	AC092634	AC055828	
VERSION	AC092634.3	GI:15809167	
KEYWORDS	HTG.		
SOURCE	Homo sapiens		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 194142)		
AUTHORS	Sulston,J.E. and Waterston,R.		
TITLE	Toward a complete human genome sequence		
JOURNAL	Genome Res. 8 (11), 1097-1108 (1998)		
MEDLINE	95063792		
PUBMED	9847074		
REFERENCE	2 (bases 1 to 194142)		
AUTHORS	Bledick,L., Haakenson,W., Dignan,G., Elliott,G. and Holmes,A.		
TITLE	The sequence of Homo sapiens BAC clone RP11-34016		
JOURNAL	Unpublished (2001)		
REFERENCE	3 (bases 1 to 194142)		
AUTHORS	Waterston,R.H.		
TITLE	Direct Submission		
JOURNAL	Submitted (19-JUL-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA		
REFERENCE	4 (bases 1 to 194142)		
AUTHORS	Waterston,R.H.		
TITLE	Direct Submission		
JOURNAL	Submitted (30-SEP-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA		
REFERENCE	5 (bases 1 to 194142)		
AUTHORS	Waterston,R.H.		
TITLE	Direct Submission		
JOURNAL	Submitted (12-OCT-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA		
REFERENCE	6 (bases 1 to 194142)		
AUTHORS	Waterston,R.		
TITLE	Direct Submission		
JOURNAL	Submitted (01-MAR-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA		
COMMENT	On Sep 30, 2001 this sequence version replaced gi:15624959. ----- Genome Center Center: Washington University Genome Sequencing Center Center code: WUGSC Web site: http://genome.wustl.edu/gsc Contact: sapience@waton.wustl.edu ----- Summary Statistics -----		

Center project name: h\_NHU340106  
Drafting Center: WIBR  
-----  
NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

NOTE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

### MAPPING INFORMATION:

**MAPPING INFORMATION:**  
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GTB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

**SOURCE INFORMATION:**

The RPl-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J. and de Jong, P. J. (1998). An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

VECTOR: pBACs3.6

## NEIGHBORING SEQUENCE INFORMATION:

RELIGIONING SEQUENCE INFORMATION:  
The clone sequenced to the left is CTD-2526L21; the clone sequenced to the right is RP11-3N2. Actual start of this clone is at base position 1 of RP11-340I6; actual end is at base position 194112 of RP11-340I6;.

The sequence of AC055828 has been incorporated into AC092634.

## FEATURES

location/Qualifiers	
1. .194142	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/chromosome="7"	
/map="7"	
/clone="RP11-34016"	
/clone.lib="RPC1-11"	
83. .392	
/rpt_family="Alu"	
399. .726	
/rpt_family="L1"	
735. .773	
/rpt_family="(TCCC)n"	
852. .1169	
/rpt_family="Alu"	
965. .1334	
/note="similar to EST BE147181 (NID:g8609905)"	
1147. .1176	
/rpt_family="(CAAAAA)n"	
1310. .1965	
/rpt_family="L1"	
2030. .2062	
/rpt_family="(CAPATA)n"	
2090. .2321	
/rpt_family="L1"	
2545. .2818	
/rpt_family="L1"	
2678. .2859	
/note="similar to EST AW992032 (NID:g8252112)"	
2829. .2874	
/rpt_family="CT-rich"	
2902. .2955	
/rpt_family="AT-rich"	
2932. .2997	
/rpt_family="L1"	
2998. .3472	
/note="similar to EST AW992032 (NID:g8252112)"	
3236. .3669	
/note="similar to EST AW753953 (NID:g7668885)"	
3689. .3732	
/rpt_family="(TG)n"	
3844. .4141	
/rpt_family="Alu"	
3852. .4242	
/note="similar to EST AV760597 (NID:q10918445)"	

misc_feature	4144..4242	/note="similar to EST AA609221 (NID:g2457649) afl2g08.sl"
misc_feature	4358..4474	/note="similar to EST AV760597 (NID:g10918445)"
misc_feature	4258..4466	/note="similar to EST AA609221 (NID:g2457649) afl2g08.sl"
repeat_region	4547..4691	/rpt_family="Alu"
repeat_region	4748..5176	/rpt_family="L1"
repeat_region	5361..5394	/rpt_family="(T)n"
repeat_region	5372..5674	/rpt_family="Alu"
repeat_region	5726..6122	/rpt_family="MaLR"
repeat_region	6125..6294	/rpt_family="Alu"
repeat_region	6302..6820	/rpt_family="(TA)n"
repeat_region	6825..6951	/rpt_family="Alu"
repeat_region	6982..7292	/rpt_family="Alu"
repeat_region	7266..7315	/rpt_family="(TAAA)n"
repeat_region	7670..7964	/rpt_family="Alu"
repeat_region	7967..8030	/rpt_family="MER1_type"
repeat_region	8278..8356	/rpt_family="(TCTA)n"
repeat_region	8332..8394	/rpt_family="ERV1"
repeat_region	8429..8730	/rpt_family="Alu"
repeat_region	9010..9514	/rpt_family="MaLR"
repeat_region	9515..9809	/rpt_family="Alu"
repeat_region	9811..9832	/rpt_family="AT_rich"
repeat_region	9816..10107	/rpt_family="Alu"
repeat_region	10108..10144	/rpt_family="MaLR"
repeat_region	10509..10864	/rpt_family="MaLR"
repeat_region	10871..11071	/rpt_family="L2"
repeat_region	11332..11643	/rpt_family="Alu"
repeat_region	12086..12243	/rpt_family="MER103"
repeat_region	12822..13094	/rpt_family="Alu"
repeat_region	13103..13292	/rpt_family="Alu"
repeat_region	14143..14300	/rpt_family="L2"
repeat_region	14301..14603	

	Query Match	8.7%	Score 87.4;	DB 9;	Length 194142;
	Best Local Similarity	67.5%;	Pred. No. 6.5e-15;		
	Matches 15%;	Conservative	0;	Mismatches 66;	Indels 9; Gaps 2;
QY	771	CACGGAGTGTATAAAGAGAGAGATTCTGTGTCGGGAATCTCCTTTGGTGGATCATCAGGGAG	830		
Dd	78694	CGTGAGTGGAGGAAGATAGGGTTCTGTGCAGGAACTCCCTTGGTGGATCATGGGGTAA	78753		
QY	831	GTGAAGCTTTTGTCAATAGCCTCATATCCAGCTTGTGTGATACCAAATTCACGTGAAGCTGG	890		
Dd	78754	GTCATGCCCTGTGGPCTGACATAAAGCCTGCTGTCTGGAGCAGACAGATCCAGTGGGGGTGG	78813		



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 24, 2002, 16:24:42 ; Search time 1812.02 Seconds  
(without alignments)  
16077.036 Million cell updates/sec

Title: US-09-708-724A-3\_COPY\_10000\_11000  
Perfect score: 1001  
Sequence: 1 caaaatttcagtttaggaaga.....gcagcacataatgtatcatg 1001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: | 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_em.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	990	98.9	151049	9	AC018558	Homo sapi
2	990	98.9	191768	9	AC092357	Homo sapi
3	631.6	63.1	144577	9	AC023824	Homo sapi
4	631.6	63.1	201061	9	AC007353	Homo sapi
5	601	60.0	191014	9	AC113268	Papio cyn
6	578.8	57.8	167133	2	AC026186	Homo sapi
7	555	55.4	167133	2	AC026186	Homo sapi
8	371.6	37.1	77060	2	AC025998	Homo sapi
9	365.2	36.5	77060	2	AC025998	Homo sapi
10	234.4	23.4	197915	9	AC005725	Homo sapi
11	101.4	10.1	193259	9	AC025281	Homo sapi
12	97.6	9.8	214269	9	AC016716	Homo sapi
13	96.2	9.6	112687	9	AC008632	Homo sapi
14	96.2	9.6	195470	2	AC022259	Homo sapi
15	95.2	9.5	147330	9	AC093514	Homo sapi
16	95.2	9.5	179636	9	AC026462	Homo sapi
17	94.6	9.5	167227	9	AC020602	Homo sapi
18	92.4	9.2	127025	9	AC018880	Homo sapi
19	92.4	9.2	203407	2	AC006174	Homo sapi
20	90.4	9.0	173719	2	AC108826	Homo sapi
21	90.4	9.0	176223	9	AC009021	Homo sapi
22	90.4	9.0	192609	9	AC009150	Homo sapi
23	90.4	9.0	208884	2	AC019092	Homo sapi
24	90.2	9.0	31994	9	AC011528	Homo sapi
25	90.2	9.0	167791	9	AC108075	Homo sapi
26	89.6	9.0	154158	2	AL136527	Homo sapi
27	89.6	9.0	191824	9	AL161935	Human DNA
28	89.4	8.9	114800	9	AC123567	Homo sapi
29	89.4	8.9	143618	2	AC026855	Homo sapi
30	89.4	8.9	162808	2	AC021340	Homo sapi
31	89.4	8.9	168872	9	AC073325	Homo sapi
32	89.4	8.9	199451	2	AC027045	Homo sapi
33	89.4	8.9	210719	2	AC087646	Homo sapi
34	89.4	8.9	213359	2	AC069006	Homo sapi
35	89.4	8.9	214738	30	AC023339	AL583802 Human DNA
36	89.2	8.9	100226	9	AL583802	Human DNA
37	89.2	8.9	172420	9	AC022182	Homo sapi
38	89.2	8.9	180541	2	AC016521	Homo sapi
39	88.4	8.8	135940	2	AC024294	Homo sapi
40	88.4	8.8	158095	2	AC016224	Homo sapi
41	88.4	8.8	160615	9	AC112236	Homo sapi
42	88.4	8.8	163974	9	AB011399	Homo sapi
43	88.4	8.8	169184	9	AC022398	Homo sapi
44	88.4	8.8	199385	2	AC019049	Homo sapi
45	88.4	8.8	201088	9	AL589733	Human DNA

ALIGNMENTS

RESULT 1  
AC018558

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AC018558 151049 bp DNA linear PRI 02-NOV-2001  
Homo sapiens chromosome 16 clone RP11-80F22, complete sequence.

AC018558

AC018558.5 GI:16596530

HTG.

Homo sapiens.

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 151049)

DOE Joint Genome Institute.

Sequencing of Human Chromosome 16

Unpublished

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

2 (bases 1 to 151049)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (14-DEC-1999) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

3 (bases 1 to 151049)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (02-NOV-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

On Nov 2, 2001 this sequence version replaced gi:9795566.  
Sequence Quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the  
GenBank flat file format but are available as part  
of this entry's ASN.1 file.  
-----  
Sequence Quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the  
GenBank flat file format but are available as part  
of this entry's ASN.1 file.  
-----

FEATURES  
source

Location/Qualifiers  
1..151049  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="RP11-80F22"  
BASE COUNT 46827 a 30783 c 30749 g 42690 t  
ORIGIN

Query Match 98.9%; Score 990; DB 9; Length 151049;  
Best Local Similarity 99.9%; Pred. No. 6.6e-231;  
Matches 1001; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 CAAAATTTTCAGTTAGGAAGAATAAGTGCAGAGATCTATTGTACTTTGGTACTACAGTTA 60  
|||||  
Db 80220 CAAAATTTTCAGTTAGGAAGAATAAGTGCAGAGATCTATTGTACTTTGGTACTACAGTTA 80279  
|||||

QY 61 ATGTATTGTCTTTCAGTAATACAGTAGATTTCCAGTGTCTTCACAAACAAATGAT 120  
|||||  
Db 80280 ATGTATTGTCTTTCAGTAATACAGTAGATTTCCAGTGTCTTCACAAACAAATGAT 80339  
|||||

QY 121 GGGTATGTGAGTAAATGATATGCAAACTAGCTTGGTTAACTTCCACATATGTG 180  
|||||  
Db 80340 GGGTATGTGAGTAAATGATATGCAAACTAGCTTGGTTAACTTCCACATATGTG 80399  
|||||

QY 181 TATTTCAAACAGTACCAATAATTCAGACAATTTTGTGTCAGTTACAATCAAAAAGTTT 240  
|||||  
Db 80400 TATTTCAAACAGTACCAATAATTCAGACAATTTTGTGTCAGTTACAATCAAAAAGTTT 80459  
|||||

QY 241 TAAATGAGACCTTAGGTGGTCTTAATCCAAATTAAGTATGATGTCTCCATGAAGAG 300  
|||||  
Db 80460 TAAATGAGACCTTAGGTGGTCTTAATCCAAATTAAGTATGATGTCTCCATGAAGAG 80519  
|||||

QY 301 AAATAAGGATACAAATGTCACACAGAGAGAAATGCCACATGAGGACACAAATGGAAGT 360  
|||||  
Db 80520 AAATAAGGATACAAATGTCACACAGAGAGAAATGCCACATGAGGACACAAATGGAAGT 80579  
|||||

QY 361 TGGCTACTTTACAAGCCTTAGGAGAGGCGCTCCGAGAAAAACACACCTTACCCACCTTGA 420  
|||||  
Db 80580 TGGCTACTTTACAAGCCTTAGGAGAGGCGCTCCGAGAAAAACACACCTTACCCACCTTGA 80639  
|||||

QY 421 TGTGTGACTTTCCTCTGTAGACGAAGTCTCCACCCTCTTCATCAGGTGGAAGCCTTTG 480  
|||||

Db 80640 TGTGTGACTTTCCTCTGTAGACGAAGTCTCCACCCTCTTCATCAGGTGGAAGCCTTTG 80699  
|||||

QY 481 ATTCTGAATATTTCTCAATGCTGGAAGGTACAAAAGTGAAGAGACAGCACAGACTCTAG 540  
|||||

Db 80700 ATTCTGAATATTTCTCAATGCTGGAAGGTACAAAAGTGAAGAGACAGCACAGACTCTAG 80759  
|||||

QY 541 GGTGAAAAGTTTAAAGAGAATAACATCTTTCCATTGCTGTGTCCTATCCCTTACACAC 600  
|||||

Db 80760 GGTGAAAAGTTTAAAGAGAATAACATCTTTCCATTGCTGTGTCCTATCCCTTACACAC 80819  
|||||

QY 601 CTATTCCAGTCTTTATTTGCTTTTGTGTTTCTGCTGCTCTGGGTATAGTGTAGTTGTA 660  
|||||

Db 80820 CTATTCCAGTCTTTATTTGCTTTTGTGTTTCTGCTGCTCTGGGTATAGTGTAGTTGTA 80879  
|||||

QY 661 ATCTCTGTTTACATACAGGATAACATAAAACAAAGTAAACATAAAATAAAACACAGACA 720  
|||||

Db 80880 ATCTCTGTTTACATACAGGATAACATAAAACAAAGTAAACATAAAATAAAACACAGACA 80939  
|||||

QY 721 GCAAACTCACTAATAGTGTGTTGGGCATGTTGACAGTGAAGACAGGAGAGTCAATAAA 780  
|||||

Db 80940 GCAAACTCACTAATAGTGTGTTGGGCATGTTGACAGTGAAGACAGGAGAGTCAATAAA 80999  
|||||

QY 781 AATGAGGTGGAAC-TTTTGAGCTAAATCAATGCTCTGTTGTTCTGTTCTTCTACAT 839  
|||||

Db 81000 AATGAGGTGGAAC-TTTTGAGCTAAATCAATGCTCTGTTGTTCTGTTCTTCTACAT 81059  
|||||

QY 840 CAGACTCTATAGTGCATGTTGTCAGTGTAGGTTTATTTATCTCTGCTCAGTAAGTG 899  
|||||

Db 81060 CAGACTCTATAGTGCATGTTGTCAGTGTAGGTTTATTTATCTCTGCTCAGTAAGTG 81119  
|||||

QY 900 CCAGAGGAGATTTTCTAACTGGGTGAGGAACAGTAGAGAGTGAAGTGAAGTGAAGACA 959  
|||||

Db 81120 CCAGAGGAGATTTTCTAACTGGGTGAGGAACAGTAGAGAGTGAAGTGAAGACA 81179  
|||||

QY 960 TTCCCTGCCATTTGCCAAAGTGGCAGCACATAATGTATCATG 1001  
|||||

Db 81180 TTCCCTGCCATTTGCCAAAGTGGCAGCACATAATGTATCATG 81221  
|||||

RESULT 2  
AC092357 191768 bp DNA linear PRI 15-AUG-2001  
LOCUS Homo sapiens chromosome 16 clone RP11-332P24, complete sequence.  
DEFINITION  
AC092357 AC0923764  
AC092357.2 GI:15187270  
VERSION  
HTG  
KEYWORDS  
Homo sapiens.  
SOURCE  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1. (bases 1 to 191768)  
DOE Joint Genome Institute.  
TITLE  
Sequencing of Human Chromosome 16  
JOURNAL  
Unpublished  
REFERENCE  
2. (bases 1 to 191768)  
DOE Joint Genome Institute.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
REFERENCE  
3. (bases 1 to 191768)  
DOE Joint Genome Institute.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (15-AUG-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
COMMENT  
On or before Aug 15, 2001 this sequence version replaced  
gi:7596817, gi:14589546.  
Sequence Quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

## Sequence Quality Assessment:

**Sequence Quality Assessment:**  
This entry has been annotated with sequence quality

estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero.

Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

**FEATURES**

source

Location/Qualifiers

1. .191768

89/T6T: T  
organism="Homo sapiens"

/db vref="tavor:0" /ofganish="HOMO

```
ab_xref="taxon:chromosome="16"
```

```
/chromosome="16"
/clone="BP11-333D3A"
```

```

/clone="RP11-332P24"
39592 C 40150 a 54877 +

```

**BASE CO  
ORIGIN**

	Query Match	98.9%	Score 990;	DB 9;	Length 191768;
	Best Local Similarity	99.9%;	Pred. No. 6.5e-231;		
	Matches 1001;	Conservative 0;	Mismatches 0;	Indels 1;	Gaps 1;
QY	1	CAAAATTTTCAGTTAGGAAAGAAATAGTCACAGAGATCTATTGTACTTGGTGACTACAGTTA	60		
Db	29965	CAAAATTTTCAGTTAGGAAAGAAATAGTCACAGAGATCTATTGTACTTGGTGACTACAGTTA	30024		
QY	61	ATGTATTGTCTCTCTTGACTAATACAGTAGAATTTTCGAGTGTCTCCACACAAAAACATGAT	120		
Db	30025	ATGTATTGTCTCTCTTGACTAATACAGTAGAATTTTCGAGTGTCTCCACACAAAAACATGAT	30084		
QY	121	GGGTATGTGAGGTAATGCATATGCAAACTAGCTTGGGTTAAACCATTCACCAATATGTGTG	180		
Db	30085	GGGTATGTGAGGTAATGCATATGCAAACTAGCTTGGGTTAAACCATTCACCAATATGTGTG	30144		
QY	181	TATTTCAAAACAGTACCATAAATGCAGACAATTTTGTGCAGTTACAATCAAAAAGTTT	240		
Db	30145	TATTTCAAAACAGTACCATAAATGCAGACAATTTTGTGCAGTTACAATCAAAAAGTTT	30204		
QY	241	TAAATGAGGACCTTAGGGTGGTCTTAATCCAACTAAAGTGATGTCTCCATGAAGAGG	300		
Db	30205	TAAATGAGGACCTTAGGGTGGTCTTAATCCAACTAAAGTGATGTCTCCATGAAGAGG	30264		
QY	301	AAATAGGATTACAAATCTGCACACAGAGAGAAATGGCCACATGAGGACACAATGAGAATG	360		
Db	30265	AAATAGGATTACAAATCTGCACACAGAGAGAAATGGCCACATGAGGACACAATGAGAATG	30324		
QY	361	TGGCTACTTTAAAGCCCTAGGAGAGAGGCCCTCCGAGAAAAACACACCCCTACCCACACCTTGA	420		
Db	30325	TGGCTACTTTAAAGCCCTAGGAGAGAGGCCCTCCGAGAAAAACACACCCCTACCCACACCTTGA	30384		
QY	421	TGTTGGACTTCATCCTGTGACAGAAAGTCCCTCCACCCCTCCTTCATCAGGTGGAAGCCCTTG	480		
Db	30385	TGTTGGACTTCATCCTGTGACAGAAAGTCCCTCCACCCCTCCTTCATCAGGTGGAAGCCCTTG	30444		
QY	481	ATTCTGAATATTCTCCAAATGCTTGGAAAGGTACAAAAGTGAAGACAGACAGACCTTCAG	540		
Db	30445	ATTCTGAATATTCTCCAAATGCTTGGAAAGGTACAAAAGTGAAGACAGACAGACCTTCAG	30504		
QY	541	GGTGAAGAGTTTAAAGAGATAACATCTTTTCCATTTGCTGTGCTTATCCCTTACACACAC	600		
Db	30505	GGTGAAGAGTTTAAAGAGATAACATCTTTTCCATTTGCTGTGCTTATCCCTTACACACAC	30564		
QY	601	CTATTCAGGCTTTATTGGTCTTTTGTGTTTTTGGTGTCTTGGGTATAGTGTAGTTGTGTA	660		
Db	30565	CTATTCAGGCTTTATTGGTCTTTTGTGTTTTTGGTGTCTTGGGTATAGTGTAGTTGTGTA	30624		
QY	661	ATCTGTGTTTACATACAGGATAACATATAAAACAAAGGTAAACAATATAAATAAAAAACACA	720		

FEATURES  
source

Location/Qualifiers  
1. 144577  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="RP11-244B22"  
BASE COUNT 39644 a 31376 c 31688 g 41869 t  
ORIGIN

Query Match 63.1%; Score 631.6; DB 9; Length 144577;  
Best Local Similarity 80.6%; Pred. No. 1.3e-143;  
Matches 830; Conservative 0; Mismatches 154; Indels 46; Gaps 6;

QY 10 AGTTAGGAAGATAAGTCAAGAGAGTCTATTGTACTTGGTACTACAGTTAATGTATTCT 69  
DB 29512 AGACAGGAGAAAGAGCATCAGATCTATTGTCATTGGTACTACAGTTAATGTATTCT 29453  
QY 70 GTTCTTGA-----CTAATACAGTAGATTTCGAGTGTCTTCTCAACAAAAACATGATGG 123  
DB 29452 GTTCTTGAACATGCTAACACGTAGATTGTGAGTCTCTCAACAGAAAAATGATGG 29393  
QY 124 TATGTGAGTAATGATATGAAACTAGCTTGGTTAAACATCCCAATATGTTGTAT 183  
DB 29392 TATGTGAGTAATGATATGAAACTAGCTTGGTTAAACATCCCAATATGTTGTAT 29333  
QY 184 TTCAAAACAGTACCAATAAATGAGACAATTTT-----GTGTCAGTTACAATCAAAAAG 237  
DB 29332 TTAAAAACAGTCCCAAAATGTAGACAATTTTATTAGTTACTAGTTACAATAAAAATG 29273  
QY 238 TTTTAAATCAGGACCTTAGGTGGTCTTAATCCAAATCTAAGTGTATGTTCCATGAAAG 297  
DB 29272 TTTTAAATCAGTACATTAGGTGGTGGCCGTGATCCAAATCTAAGTGTATGTTCCATGAAAG 29213  
QY 298 AGCAATATAGGATACAAATGTGCACACAGAGAAATGCCACATGAGGACAAATGACA 357  
DB 29212 AGGAGATAGGATACAAATGTGCACACAGAGAAATGCCACATGAGGACAAATGACA 29153  
QY 358 ATGTGCTACTTACAAGCTTAGGAGAGGCGCTCCAGAGAAAACACACCTTACCACACCT 417  
DB 29152 AGGTGCCACTTACACACCTTAGGAGAGAGCGCTTGGGGCAGACACACCTTGGCCACCT 29093  
QY 418 TGATGTTTGA-----CTTATCTCTGTAGAGAGAGTCTCCAACTCTGAGGATCAAAA 516  
DB 29032 TCCTTCATCATGTTGAACCTTCTGACCCCTGAATACCTCTCAATGCTGTAATGTACCAAG 28973  
QY 517 GTGAAGAGACAGACAGACCTCAGGTTGAAAAGTTTAAAGAGAAATTAACATCTTTCCATTG 576  
DB 28972 GTGAAGGACAGACAGACCTCAGGTTGAAAAGTTTAAAGAGAAATTAACATCTTTCCATTG 28913  
QY 577 CTGTGCTTATCCCTTACACACACCTTATCCAGTCTTTTATGTTGCTTTTGTGTTTCGGT 636  
DB 28912 CCCTGTCCATCCCAACACACACCTTGTCCAGCCCTTATGTTGCTTTTGTATTCCTCTG 28853  
QY 637 TCCTGGGTATAGTGTAGTTGTTAAATCTGTGTTATACATACAGGATACATAAACAAGG 696  
DB 28852 TCCTGGATATAGGTTAACTTATTAATCTGTGTTATGTATGATAGGATACATAAACAAGG 28793  
QY 697 TAAACAATAAATAAACAAGACAGCAAACTCAACTAATAGTTTGGGATGTTGACA 756  
DB 28792 TAAACAATAAATAAATAAACAAGCAAACTCAACTAATAGTTTGGGATGTTGACA 28733  
QY 757 GTGAAGACAGGAGTACATAAATAAATGGAGTGGAACT--TTTGTAGCTAAATCAATGTCC 815  
DB 28732 GTGAAGGACAGGAGTACATAAATAAATGGAGTGGAACTTCTTGGAGCTCAGTCCACATCC 28673  
QY 816 TGTGTGTTCTGTTTCTACATCAGACTCTATAGTGGCAATGTACAGTTAGTGTGTT 875  
DB 28672 TGT-----TGTGTTTCTACATCAGAACTATATAGTACAGTATTTCAGGTTTGTGTT 28621

QY 876 TTATCTTCCCTCAGTAACTGCGAGAGAGATTTTCTAACTGGGTGAGAACAGG 935  
DB 28620 TTGTCTTCCCTGCGCCCAATAAGTCCAGAGGGATTTATCTAATCAGGTGAGGACTGG 28561  
QY 936 TA----GAGAACTGTAAGTGAGACAAACTTCCCTGCCAFTTCCCAAAAGTGGCAGCACATA 991  
DB 28560 TAACTTGCACTTTTGAAGTGAAGACAAATCTTCTTGCATTTGCCACAGTGGCATGGCAGA 28501  
QY 992 ATGTATCATG 1001  
DB 28500 ATTTATCATG 28491

RESULT 4  
AC007353/c

LOCUS AC007353 201061 bp DNA linear PRI 01-JUN-2001  
DEFINITION Homo sapiens chromosome 16 clone RP11-488I20, complete sequence.  
ACCESSION AC007353  
VERSION AC007353.4 GI:14277247  
KEYWORDS HTG  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 201061)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Human Chromosome 16  
JOURNAL Unpublished  
AUTHORS

REFERENCE 2 (bases 1 to 201061)  
Bruce, D., Mundt, M., Doggett, N., Munk, C., Saunders, E., Robinson, D.,  
Jones, M., Buckingham, J., Chasteen, L., Thompson, S., Goodwin, L.,  
Bryant, J., Tesmer, J., Meincke, L., Longmire, J., White, S., Tatum, O.,  
Campbell, C., Fawcett, J., Maltbie, M., Bussod, M., Sutherland, R.,  
McMurry, K., Han, C. and Deaven, L.  
TITLE Direct Submission  
JOURNAL Submitted (22-APR-1999) Center for Human Genome Studies, DOE Joint  
Genome Institute, Los Alamos National Laboratory, MS M888, Los  
Alamos, NM 87545, USA

REFERENCE 3 (bases 1 to 201061)  
DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (01-JUN-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On Jun 1, 2001 this sequence version replaced gi:9795555.  
COMMENT -----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: <http://www.jgi.doe.gov>

-----  
Sequence Quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the  
GenBank flat file format but are available as part  
of this entry's ASN.1 file.

-----  
Location/Qualifiers  
1. 201061  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="RP11-488I20"  
BASE COUNT 57062 a 41994 c 41012 g 60993 t  
ORIGIN

FEATURES  
source

Query Match 63.1%; Score 631.6; DB 9; Length 201061;  
Best Local Similarity 80.6%; Pred. No. 1.3e-143;  
Matches 830; Conservative 0; Mismatches 154; Indels 46; Gaps 6;





QY	302	AATAAGGATACAAATGTGCACACAGAGAGAAATGGCCACATGAGGACACAAATGAGAATGT	361
Db	190575	GATAAAGATGCAAAATGTGCACACTTGGAGAAATGGCCAGTGGAGGATAAAGTAGAAGGC	190516
QY	362	GGCTACTTACAAGCTAGAGAGAGAGCCCTCCGAGAAAACACACACCCTACCACACCTTGAT	421
Db	190515	AGCCACTTAACCTCAGGAGAGAGCCCTGGAGCAAAACCCCTGCCACATCTTGAT	190456
QY	422	GTGGACTTCATCTCTAGACGAAGTCCCTCCACCCCTTCATCAGGTGGAGCCCTTGA	481
Db	190455	CTTGACTTAT-----TCCTCTCATCAAGTGGAGCTTCTGA	190417
QY	482	TTCTGAATATTCCTCAAAATGCTGGAAGGTACAAAAGTGAAGACAGACACACCTTCAG	541
Db	190416	TCCTGAATACTTTCCAAAAACTAGAAGCTACAATGCTGAGGGGCGACAGACCTTCAGS	190357
QY	542	GTGAAAAGTTTAAAGAGAATAACATCTTCCATTTGCTGTGCTATCCCTACACACACC	601
Db	190356	GTGAAAAGTTTAAAGAGGATAACACATCCCATTTGCCCTATCCTATCCCAATGCACACC	190297
QY	602	TATTCAGTCTTATTTGGTCTTTTGTGTTTTCGTGCTCCCTGGGTATAGTGTAGTTCTGAAA	661
Db	190296	TGTGCCAGCCTTATTTGGTGTGTTTGTATTTCCCTGTACTGGGGATAGTGTAGCTTGTAA	190237
QY	662	TCGTGCTTTACATACAGGATAACATAAAACAAAGGTAAACAATAAATAAAACACAGAC	721
Db	190236	TTTGTGTTTACATAGAGGATAACATAAAACAAAGGTAAACAATAAATAAAACAAACAG	190177
QY	722	CAAACTCAACTAATAGTGTGGGATGTTGGGATGGTGACAGTGAAGACAGAGAGTGCATAAA	781
Db	190176	CACAATCAAACTACTAGTGTGGGAGGGGTGACAGTGAAGGAAGGAATGTCTATGTAATA	190117
QY	782	ATGAGGTGGAAC-TTTTGAGCTAAATCAATGCTCTGTTGTTGTTCTGTCTACATC	840
Db	190116	ATGGAATGAGGCTTTTGTAGCTAAATCAATGCTC-----TTTGTGTTTCTACATC	190065
QY	841	AGACTATAGTGGCAATGTTCAGGTAGGTGTTTATTCCTTGGCTTCAGTCAAGTGC	900
Db	190064	AGAATCTGTAGTACAAATGTTCAGGTAGGTGTTTGTGCTTATTCGCTCATAAGTGC	190005
QY	901	CAGAGAGATTTTCTAAACTGGGTGAGGACAGGTAGAGAGTGAAGTGACAGAACT	960
Db	190004	CAGAGGGATTAATCTCTAAACTGGGTGAGAAACAAGTCAAGAGTGAAGTGACAGAAAGT	189945
QY	961	TCCTCGCAATTTGCCAAAGTGCCAGCACATAATGTATCATG	1001
Db	189944	TCCTCGGATTTGCCACAGTGGCAGGAGAAATTCATCATG	189904
RESULT	6		
AC026186/c			
LOCUS		167133 bp	DNA linear HTG 29-MAY-2000
DEFINITION		Homo sapiens chromosome 3 clone RP11-303P8 map 3p, WORKING DRAFT	
SEQUENCES		44 unordered pieces.	
AC026186			
VERSION		AC026186.2	GI:8101183
KEYWORDS		HTG; HTGS_PHASE1; HTGS_DRAFT.	
SOURCE		Homo sapiens.	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS		1 (bases 1 to 167133)	
		Bao, J., Fan, H., Peng, X., Guan, Q., Gu, X., Guo, D., He, L., Hu, S., Dong, W., Fan, H., Peng, X., Wang, J., Zhang, Y., Zhang, H., Liu, B., Huang, F., Jin, Y., Kang, N., Li, C., Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, B., Liu, N., Liu, B., Li, W., Li, W., Li, Y., Luo, J., Niu, Y., Qi, Q., Qi, X., Song, S., Sun, M., Sun, W., Sun, Y., Tao, R., Wang, H., Wang, J., Wang, L., Wang, L., Wang, R., Wang, X., Wang, X., Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, M., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Yu, J. and Yang, H.	
Chromosome 3p genomic sequence			
Unpublished			
2 (bases 1 to 167133)			
Chen, J., Hu, S., Dong, W., Wang, J., Zhang, Y., Zhang, H., Liu, B., Bao, W., Sun, Y., Wu, Q., Wang, H., Yang, X., Cheng, C., Wang, Y., Niu, Y., Qi, X., Li, T., Zhang, H., Liu, N., Wu, D., Yu, B., Fan, H., Liu, Y., Li, G., Li, C., Bao, Q., Bao, J., Wang, X., Song, L., Zhang, L., Guo, D., Huang, F., Zhang, G., Li, J., Bian, X., Zhang, M., Li, L., Feng, X., Yu, J. and Yang, H.			
Direct Submission			
Submitted (21-MAR-2000) Human Genomic Center, Institute of Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing 100101, P.R.China			
On May 29, 2000 this sequence version replaced gi:7272003.			
-----Genome Center			
Center:Beijing Center			
Center code:Beijing			
Website:http://hgclgtp.ac.cn			
http://www.genomics.org.cn			
Contact:hgclgtp.ac.cn			
----- Project Information			
Center project name:1% project			
Center clone name: RP11-303P8			
----- Summary Statistics			
Sequencing vector: pUC18; 100% of reads			
Chemistry: Dye-terminator: ET 55% of reads			
Chemistry: Dye-terminator: Big Dye; 45% of reads			
Assembly program: Phrap; version 0.990329			
Consensus quality: 144681 bases at least Q40			
Consensus quality: 152516 bases at least Q30			
Consensus quality: 159163 bases at least Q20			
Insert size: 121837; sum-of-contigs			
Quality coverage: 4.43x in Q20 bases;sum-of-contigs			
-----			
* NOTE: This is a 'working draft' sequence. It currently			
* consists of 44 contigs. The true order of the pieces			
* is not known and their order in this sequence record is			
* arbitrary. Gaps between the contigs are represented as			
* runs of N, but the exact sizes of the gaps are unknown.			
* This record will be updated with the finished sequence			
* as soon as it is available and the accession number will			
* be preserved.			
-----			
1 2054: contig of 2054 bp in length			
2055 2154: gap of unknown length			
2155 3414: contig of 1260 bp in length			
3415 3514: gap of unknown length			
3515 5281: contig of 1767 bp in length			
5282 5381: gap of unknown length			
5382 6678: contig of 1297 bp in length			
6679 6778: gap of unknown length			
6779 8533: contig of 1755 bp in length			
8534 8633: gap of unknown length			
8634 9889: contig of 1356 bp in length			
9890 10089: gap of unknown length			
10090 11298: contig of 1209 bp in length			
11299 11398: gap of unknown length			
11399 12931: contig of 1533 bp in length			
12932 12932: gap of unknown length			
13032 14756: contig of 1725 bp in length			
14757 14856: gap of unknown length			
14857 16838: contig of 1982 bp in length			
16839 16938: gap of unknown length			
16939 18630: contig of 1692 bp in length			
18631 18730: gap of unknown length			
18731 20330: contig of 1600 bp in length			
20331 20430: gap of unknown length			
20431 21642: contig of 1212 bp in length			
21643 21742: gap of unknown length			
21743 23363: contig of 1621 bp in length			
23364 23463: gap of unknown length			
23464 25148: contig of 1685 bp in length			
25149 25248: gap of unknown length			
25249 27496: contig of 2248 bp in length			

```
* 27497 27596: gap of unknown length
* 27597 30050: contig of 2454 bp in length
* 30051 30150: gap of unknown length
* 30151 32351: contig of 2201 bp in length
* 32352 32451: gap of unknown length
* 32452 34424: contig of 1973 bp in length
* 34425 34524: gap of unknown length
* 34525 36830: contig of 2306 bp in length
* 36831 36930: gap of unknown length
* 36931 39686: contig of 2756 bp in length
* 39687 39786: gap of unknown length
* 39787 42512: contig of 2726 bp in length
* 42513 42612: gap of unknown length
* 42613 44556: contig of 1944 bp in length
* 44557 44656: gap of unknown length
* 44657 48395: contig of 3739 bp in length
* 48396 48495: gap of unknown length
* 48496 50935: contig of 2440 bp in length
* 50936 51035: gap of unknown length
* 51036 54435: contig of 3400 bp in length
* 54436 54535: gap of unknown length
* 54536 59749: contig of 5214 bp in length
* 59750 59849: gap of unknown length
* 59850 63021: contig of 3172 bp in length
* 63022 63121: gap of unknown length
* 63122 66858: contig of 3737 bp in length
* 66859 71348: gap of unknown length
* 71349 71448: gap of unknown length
* 71449 75783: contig of 4335 bp in length
* 75784 75883: gap of unknown length
* 75884 79500: contig of 3617 bp in length
* 79501 79600: gap of unknown length
* 79601 83516: contig of 3916 bp in length
* 83517 83616: gap of unknown length
* 83617 88942: contig of 5326 bp in length
* 88943 89042: gap of unknown length
* 89043 95244: contig of 6202 bp in length
* 95245 95344: gap of unknown length
* 95345 99266: contig of 3922 bp in length
* 99267 99366: gap of unknown length
* 99367 104303: contig of 4937 bp in length
* 104304 104403: gap of unknown length
* 104404 110029: contig of 5626 bp in length
* 110030 110129: gap of unknown length
* 110130 117586: contig of 7457 bp in length
* 117587 117686: gap of unknown length
* 117687 125138: contig of 7452 bp in length
* 125139 125238: gap of unknown length
* 125239 132290: contig of 7052 bp in length
* 132291 132390: gap of unknown length
* 132391 142718: contig of 10328 bp in length
* 142719 142818: gap of unknown length
* 142819 154592: contig of 11774 bp in length
* 154593 154692: gap of unknown length
* 154693 167133: contig of 12441 bp in length.
FEATURES
    source
        1. .167133
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="3"
            /map="3p"
            /clone="RP11-303P8"
            1. .2054
                /note="assembly_name:Contig17"
                2155. .3414
                    /note="assembly_name:Contig18"
                    3515. .5281
                        /note="assembly_name:Contig19"
                        5382. .6678
                            /note="assembly_name:Contig21"
                            6779. .8533
                                /note="assembly_name:Contig22"
```

```
misc_feature 8634. .9989
              /note="assembly_name:Contig25"
              10090. .11298
              /note="assembly_name:Contig27"
              11399. .12931
              /note="assembly_name:Contig28"
              13032. .14756
              /note="assembly_name:Contig29"
              14857. .16838
              /note="assembly_name:Contig30"
              16939. .18630
              /note="assembly_name:Contig31"
              18731. .20330
              /note="assembly_name:Contig32"
              20431. .21642
              /note="assembly_name:Contig33"
              21743. .23363
              /note="assembly_name:Contig34"
              23464. .25148
              /note="assembly_name:Contig35"
              25249. .27496
              /note="assembly_name:Contig36"
              27597. .30050
              /note="assembly_name:Contig37"
              30151. .32351
              /note="assembly_name:Contig38"
              32452. .34424
              /note="assembly_name:Contig39"
              34525. .36830
              /note="assembly_name:Contig40"
              36931. .39686
              /note="assembly_name:Contig41"
              39787. .42512
```

Query Match 57.8%; Score 578.8; DB 2; Length 167133;  
Best Local Similarity 81.0%; Pred. No. 9.7e-131;  
Matches 812; Conservative 0; Mismatches 142; Indels 48; Gaps 10;

```
QY 8 TCAGTTAGGAAGATAAGTGCAGAGATCTATTGTACTTTGGTGACTACAGTTAATGTATT 67
Db 137575 TTAGACAGGAGGAATAAGTGCAGAGATCTATTGTACTTTGGTGACTACAGTTAATGTATT 137516

QY 68 GTGTTCTTGA-----CTAATACAGTAGATTTCGAGTGTCTCACAACAAAAACATGATG 121
Db 137515 CTGTTCTTGAACATTCGTAGGACAGTAGATTTTGTAGTATCTCACAACAAAAA-AAGATG 137457

QY 122 GGTATGTGAGGTAATGCATATG-CAAACTAGCTTGGGTTAACCATTCACATATGTGTG 180
Db 137456 GGTATGTGAGGTAATGCATATGTTTCAGTTATCTTGGGTTAACCATTCACATATGTGTGCA 137397

QY 181 TATTTCAAAACAGTACCATAAATGCAGACAAATTTGTGTCAGTTACAATCAAAAAGTTT 240
Db 137396 TATTTCAACACAGTACCATAAATGTAGGCAATTTT-TATCAGTTACA---AAAAAATTT 137341

QY 241 TAAATGAGGACCTTAGGGTGGTCTCTAATCCAATCTAATGATGTCTCCATGAAGAGG 300
Db 137340 TTAATGAGGACCTTAGGGAGAACCTTAATCCAATGTAACTGTGTCTCCACAAGAGG 137281

QY 301 AAATAAGGATACAAATGTCACACAGAGAGAAATGCCACATGAGGACACATGAGAATG 360
Db 137280 AGATAAAGATACAAACGTTGCACACAGAGAAATGCCACATGAGGACATAGGAGAAGG 137221

QY 361 TGGCTACTTACAAGCCTTAGGAGAGAGGCTCCGAGAAACACACACCTACCCACACTTCA 420
Db 137220 CGGCCACTTTACAAACCCAGGAGAGAGGCTCAGAGAAAAACCCCTTGCCACACTTTGA 137161

QY 421 TGTGGACTTCATCTCTAGACGAAAGTCTCCACCTCTCTCATCAGTGGTGAAGCCTTTG 480
Db 137160 TCTTGGACTTCAT-----TCTCCTTCATCAATGGAGGCTTCTG 137122

QY 481 ATCTGAATATTTCTCCAAATGCTGGAAGGTACAAAGGTACAAAGTGAAGACAGACAGACTCAG 540
Db 137121 ATCCTGAATACTTTCCAAAACACTAGAAAAGTACAGTGGTGAAGGACGACAGAACTCAG 137062
```

QY 541 GGTGAAAGTTTAAAGAGATAACATCTTTCCATCTGCTGTGCTCCATCCCTACACAC 600  
|||||  
Db 137061 GGTGAAAGTTTAAAGAGATAACACCTCCCATTCGCCCTGCTCTATCCCAACACAC 137002  
QY 601 CTAATCCAGTCCTTATGTCCTTTTGTGTTTTCGTCCTCGGTATAGTGTAGTTGTTAA 660  
|||||  
Db 137001 CTGTGCCAGCCCTTATGTTGTTTGTGCTATCCCTGCTCTCGGGGATAGTGTAACTGTTG 136942  
QY 661 ATCTGTGTTTACATACAGATACATATAAACAAGGTAACAATAAATAAACAAGACACA 720  
|||||  
Db 136941 ATTGTGTTTACATAGATACGTATAAACAAGGTAACAATAAATAAACAAGACACA 136882  
QY 721 GCAAACTCAACTAATAGTGTGTTGGCATGTCACAGTGAACAGAGGAGTGCACATAAA 780  
|||||  
Db 136881 GCACAACCTCAATAATAGTGTGTTGGCGGGTGACAGTGCAGGGAAGAAG-GCCACATAC 136823  
QY 781 ATGGAGGTGGAACT-TTTTGAGCTAAATCAATGCTGTTGTTGTTCTGTTCTTACAT 839  
|||||  
Db 136822 AATGGAGATGGAGCTTTTGAACATAATCAATGCTGTTGTT-----TGTTCCTACAT 136772  
QY 840 CAGACTCTATAGTGGCAATGTCAGTGTAGGTTTATATCCTGCTGCTCAGTAAAG 899  
|||||  
Db 136771 CAGAATCTATAGTACAAATGTCAGTGTAGGTTTATATCCTGCTGCTCAGTAAAG 136712  
QY 900 CCAGAGGAGATTTTCTAAACTGGGTGAGGAACAGGTAGTGAAGTGTAAAGTGCAGACAAAC 959  
|||||  
Db 136711 CCAGAGGGGATTTCTAAACTGGGTGAGGAACAGGTCAAGGAGT---GTGAGACAAAC 136656  
QY 960 TTCCCTGCCATTTGCCAAGTGGCAGCAGACATATATGATCATG 1001  
|||||  
Db 136655 TTCCCTGCCATTTGCCAAGTGGCAGCAGACATATGATCATG 136614

## RESULT 7

AC026186 167133 bp DNA linear HTG 29-MAY-2000  
LOCUS Homo sapiens chromosome 3 clone RP11-303P8 map 3p, WORKING DRAFT  
DEFINITION SEQUENCE, 44 unordered pieces.

AC026186

AC026186.2 GI:8101183

HTG: HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE Homo sapiens.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 167133)

Bao,J., Bao,Q., Bao,W., Bian,X., Cao,T., Chen,C., Chen,J., Ding,H., Dong,W., Fan,H., Feng,X., Guan,Q., Gu,X., Guo,D., He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,B., Liu,Y., Li,W., Li,W., Li,Y., Luo,J., Niu,Y., Qi,Q., Qi,X., Song,S., Sun,M., Sun,W., Sun,Y., Tao,R., Wang,H., Wang,J., Wang,J., Wang,L., Wang,L., Wang,R., Wang,X., Wang,X., Wang,Y., Wu,D., Wu,Q., Xie,F., Xuan,Z., Xue,Y., Yan,C., Yang,X., Yu,B., Zeng,Y., Zhang,G., Zhang,H., Zhang,H., Zhang,L., Zhang,M., Zhang,X., Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu,B., Yu,J. and Yang,H.

Chromosome 3p genomic sequence

Unpublished

2 (bases 1 to 167133)

Chen,J., Hu,S., Dong,W., Wang,J., Zhang,Y., Zhang,H., Liu,B., Bao,W., Sun,Y., Wu,Q., Wang,H., Yang,X., Cheng,C., Wang,Y., Niu,Y., Qi,X., Li,T., Zhang,H., Liu,N., Wu,D., Yu,B., Fan,H., Liu,Y., Li,G., Li,C., Bao,Q., Bao,J., Wang,X., Song,L., Zhang,L., Guo,D., Huang,F., Zhang,G., Li,J., Bian,X., Zhang,M., Li,L., Feng,X., Yu,J. and Yang,H.

Direct Submission

Submitted (21-MAR-2000) Human Genomic Center, Institute of Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing 100101, P.R.China  
On May 29, 2000 this sequence version replaced gi:7272003.

-----Genome Center

Center:Beijing Center

Center code:Beijing

Website:http://hg.igtp.ac.cn

http://www.genomics.org.cn

Contact:hg@igtp.ac.cn

----- Project Information

Center project name:1# project

Center clone name: RP11-303P8

----- Summary Statistics

Sequencing vector: pUC18; 100% of reads

Chemistry: Dye-terminator: ET 5% of reads

Chemistry: Dye-terminator: Big Dye; 45% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 144681 bases at least Q40

Consensus quality: 152516 bases at least Q30

Consensus quality: 159163 bases at least Q20

Insert size: 121837; sum-of-contigs

Quality coverage: 4.43x in Q20 bases;sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently consists of 44 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 2054: contig of 2054 bp in length  
\* 2055 2154: gap of unknown length  
\* 2155 3414: contig of 1260 bp in length  
\* 3415 5281: contig of 1767 bp in length  
\* 5282 5382: gap of unknown length  
\* 5382 6678: contig of 1297 bp in length  
\* 6679 6779: gap of unknown length  
\* 6779 8533: contig of 1755 bp in length  
\* 8534 8633: gap of unknown length  
\* 8634 9989: contig of 1356 bp in length  
\* 9990 10089: gap of unknown length  
\* 10090 11298: contig of 1209 bp in length  
\* 11299 11398: gap of unknown length  
\* 11399 12931: contig of 1333 bp in length  
\* 12932 13031: gap of unknown length  
\* 13032 14756: contig of 1725 bp in length  
\* 14757 14856: gap of unknown length  
\* 14857 16838: contig of 1982 bp in length  
\* 16839 16938: gap of unknown length  
\* 16939 18630: contig of 1692 bp in length  
\* 18631 18730: gap of unknown length  
\* 18731 20330: contig of 1600 bp in length  
\* 20331 20430: gap of unknown length  
\* 20431 21642: contig of 1212 bp in length  
\* 21643 21742: gap of unknown length  
\* 21743 23363: contig of 1621 bp in length  
\* 23364 23463: gap of unknown length  
\* 23464 25148: contig of 1685 bp in length  
\* 25149 25248: gap of unknown length  
\* 25249 27496: contig of 2248 bp in length  
\* 27497 27596: gap of unknown length  
\* 27597 30050: contig of 2454 bp in length  
\* 30051 30150: gap of unknown length  
\* 30151 32351: contig of 2201 bp in length  
\* 32352 32451: gap of unknown length  
\* 32452 34424: contig of 1973 bp in length  
\* 34425 34524: gap of unknown length  
\* 34525 36830: contig of 2306 bp in length  
\* 36831 36930: gap of unknown length  
\* 36931 39686: contig of 2756 bp in length  
\* 39687 39786: gap of unknown length  
\* 39787 42612: contig of 2726 bp in length  
\* 42613 44556: contig of 1944 bp in length  
\* 44557 44656: gap of unknown length  
\* 44657 48395: contig of 3739 bp in length



```
QY 849 TAGTGCAATGTCAGGTTAGGTTTATCTTCTGCTCAGTAAGTGCAGAGGAG 908
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 102283 TAGTAGCAATGTCAGGTTACGTTGTTTGTCTTCTGCTCCTTAAGTGCAGAGGG 102342

QY 909 ATTTCCTAAATCGGTGAGGACAGGTAGAGTAAGTGCAGAAACTTCCTGCC 968
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 102343 ATTATTCCTAAATCGGTGAGGACAGGTAGAGTAAGTGCAGAAACTTCCTGCC 102402

QY 969 ATTTCGCAAGTGGCAGCACATAATGTCATCATG 1001
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 102403 ATTTCGCAAGTGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 102435

RESULT 8
AC025998/c
LOCUS 77060 bp DNA linear HTG 19-MAR-2000
DEFINITION Homo sapiens chromosome 10 clone RP11-302C6 map 10, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC025998
VERSION AC025998.1 GI:7263216
KEYWORDS HTG; HTGS_PHASE0.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 77060)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 10, clone RP11-302C6
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 77060)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collumore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Latouche,K., Lamazares,R., Landers,T., Lehoczyk,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,T., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
TITLE Submitted (19-MAR-2000) Whitehead Institute/MIT Center for Genome
JOURNAL Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8116
Center clone name: 302_C_6
-----
* NOTE: This record contains 86 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
```

```
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 773: contig of 773 bp in length
* 774 873: gap of 100 bp
* 874 1684: contig of 811 bp in length
* 1685 1784: gap of 100 bp
* 1785 2572: contig of 788 bp in length
* 2573 2672: gap of 100 bp
* 2673 3467: contig of 795 bp in length
* 3468 3567: gap of 100 bp
* 3568 4354: contig of 787 bp in length
* 4355 4454: gap of 100 bp
* 4455 5254: contig of 800 bp in length
* 5255 5354: gap of 100 bp
* 5355 6140: contig of 786 bp in length
* 6141 6240: gap of 100 bp
* 6241 7045: contig of 805 bp in length
* 7046 7145: gap of 100 bp
* 7146 7932: contig of 787 bp in length
* 7933 8032: gap of 100 bp
* 8033 8851: contig of 819 bp in length
* 8852 8951: gap of 100 bp
* 8952 9754: contig of 803 bp in length
* 9755 9854: gap of 100 bp
* 9855 10640: contig of 786 bp in length
* 10641 10740: gap of 100 bp
* 10741 11547: contig of 807 bp in length
* 11548 11647: gap of 100 bp
* 11648 12448: contig of 801 bp in length
* 12449 12548: gap of 100 bp
* 12549 13346: contig of 798 bp in length
* 13347 13446: gap of 100 bp
* 13447 14234: contig of 788 bp in length
* 14235 14334: gap of 100 bp
* 14335 15106: contig of 772 bp in length
* 15107 15206: gap of 100 bp
* 15207 16013: contig of 807 bp in length
* 16014 16113: gap of 100 bp
* 16114 16924: contig of 811 bp in length
* 16925 17024: gap of 100 bp
* 17025 17798: contig of 774 bp in length
* 17799 17898: gap of 100 bp
* 17899 18706: contig of 808 bp in length
* 18707 18806: gap of 100 bp
* 18807 19616: contig of 810 bp in length
* 19617 19716: gap of 100 bp
* 19717 20523: contig of 807 bp in length
* 20524 20623: gap of 100 bp
* 20624 21400: contig of 777 bp in length
* 21401 21500: gap of 100 bp
* 21501 22285: contig of 785 bp in length
* 22286 22385: gap of 100 bp
* 22386 23185: contig of 800 bp in length
* 23186 23285: gap of 100 bp
* 23286 24103: contig of 818 bp in length
* 24104 24203: gap of 100 bp
* 24204 25012: contig of 809 bp in length
* 25013 25112: gap of 100 bp
* 25113 25922: contig of 810 bp in length
* 25923 26022: gap of 100 bp
* 26023 26839: contig of 817 bp in length
* 26840 26939: gap of 100 bp
* 26940 27765: contig of 826 bp in length
* 27766 27865: gap of 100 bp
* 27866 28662: contig of 797 bp in length
* 28663 28762: gap of 100 bp
* 28763 29574: contig of 812 bp in length
* 29575 29674: gap of 100 bp
* 29675 30453: contig of 779 bp in length
* 30454 30553: gap of 100 bp
```



Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,  
 O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,  
 Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
 Stange-Thoman, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
 Testaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J.,  
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,  
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.

# Direct Submission

Submitted (19-MAR-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L8116

Center clone name: 302\_C\_6

-----

\* NOTE: This record contains 86 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

\* 1 773: contig of 773 bp in length  
 \* 774 873: gap of 100 bp  
 \* 874 1684: contig of 811 bp in length  
 \* 1685 1784: gap of 100 bp  
 \* 1785 2572: contig of 788 bp in length  
 \* 2573 2672: gap of 100 bp  
 \* 2673 3467: contig of 795 bp in length  
 \* 3468 3567: gap of 100 bp  
 \* 3568 4354: contig of 787 bp in length  
 \* 4355 4454: gap of 100 bp  
 \* 4455 5254: contig of 800 bp in length  
 \* 5255 5354: gap of 100 bp  
 \* 5355 6140: contig of 786 bp in length  
 \* 6141 6240: gap of 100 bp  
 \* 6241 7045: contig of 805 bp in length  
 \* 7046 7145: gap of 100 bp  
 \* 7146 7932: contig of 787 bp in length  
 \* 7933 8032: gap of 100 bp  
 \* 8033 8851: contig of 819 bp in length  
 \* 8852 8951: gap of 100 bp  
 \* 8952 9754: contig of 803 bp in length  
 \* 9755 9854: gap of 100 bp  
 \* 9855 10640: contig of 786 bp in length  
 \* 10641 10740: gap of 100 bp  
 \* 10741 11547: contig of 807 bp in length  
 \* 11548 11647: gap of 100 bp  
 \* 11648 12448: contig of 801 bp in length  
 \* 12449 12548: gap of 100 bp  
 \* 12549 13346: contig of 798 bp in length  
 \* 13347 13446: gap of 100 bp  
 \* 13447 14234: contig of 788 bp in length  
 \* 14235 14334: gap of 100 bp  
 \* 14335 15106: contig of 772 bp in length  
 \* 15107 15206: gap of 100 bp  
 \* 15207 16013: contig of 807 bp in length  
 \* 16014 16113: gap of 100 bp  
 \* 16114 16924: contig of 811 bp in length  
 \* 16925 17024: gap of 100 bp  
 \* 17025 17798: contig of 774 bp in length

\* 17799 17898: gap of 100 bp  
 \* 17899 18706: contig of 808 bp in length  
 \* 18707 18806: gap of 100 bp  
 \* 18807 19616: contig of 810 bp in length  
 \* 19617 19716: gap of 100 bp  
 \* 19717 20523: contig of 807 bp in length  
 \* 20524 20623: gap of 100 bp  
 \* 20624 21400: contig of 777 bp in length  
 \* 21401 21500: gap of 100 bp  
 \* 21501 22285: contig of 785 bp in length  
 \* 22286 22385: gap of 100 bp  
 \* 22386 23185: contig of 800 bp in length  
 \* 23186 23286: gap of 100 bp  
 \* 23286 24103: contig of 818 bp in length  
 \* 24104 24203: gap of 100 bp  
 \* 24204 25012: contig of 809 bp in length  
 \* 25013 25112: gap of 100 bp  
 \* 25113 25922: contig of 810 bp in length  
 \* 25923 26022: gap of 100 bp  
 \* 26023 26839: contig of 817 bp in length  
 \* 26840 26939: gap of 100 bp  
 \* 26940 27765: contig of 826 bp in length  
 \* 27766 27865: gap of 100 bp  
 \* 27866 28662: contig of 797 bp in length  
 \* 28663 28762: gap of 100 bp  
 \* 28763 29574: contig of 812 bp in length  
 \* 29575 29674: gap of 100 bp  
 \* 29675 30453: contig of 779 bp in length  
 \* 30454 30553: gap of 100 bp  
 \* 30554 31339: contig of 786 bp in length  
 \* 31340 31439: gap of 100 bp  
 \* 31440 32242: contig of 803 bp in length  
 \* 32243 32342: gap of 100 bp  
 \* 32343 33137: contig of 795 bp in length  
 \* 33138 33237: gap of 100 bp  
 \* 33238 34042: contig of 805 bp in length  
 \* 34043 34142: gap of 100 bp  
 \* 34143 34961: contig of 819 bp in length  
 \* 34962 35061: gap of 100 bp  
 \* 35062 35863: contig of 802 bp in length  
 \* 35864 35963: gap of 100 bp  
 \* 35964 36767: contig of 804 bp in length  
 \* 36768 36867: gap of 100 bp  
 \* 36868 37646: contig of 779 bp in length  
 \* 37647 37746: gap of 100 bp  
 \* 37747 38559: contig of 813 bp in length  
 \* 38560 38659: gap of 100 bp  
 \* 38660 39458: contig of 799 bp in length  
 \* 39459 39558: gap of 100 bp  
 \* 39559 40371: contig of 813 bp in length  
 \* 40372 40471: gap of 100 bp  
 \* 40472 41210: contig of 739 bp in length  
 \* 41211 41310: gap of 100 bp  
 \* 41311 42103: contig of 793 bp in length  
 \* 42104 42203: gap of 100 bp  
 \* 42204 43003: contig of 800 bp in length  
 \* 43004 43103: gap of 100 bp  
 \* 43104 43905: contig of 802 bp in length  
 \* 43906 44005: gap of 100 bp  
 \* 44006 44825: contig of 820 bp in length  
 \* 44826 44925: gap of 100 bp  
 \* 44926 45722: contig of 797 bp in length  
 \* 45723 45822: gap of 100 bp  
 \* 45823 46629: contig of 807 bp in length  
 \* 46630 46729: gap of 100 bp  
 \* 46730 47549: contig of 820 bp in length  
 \* 47550 47649: gap of 100 bp  
 \* 47650 48440: contig of 791 bp in length  
 \* 48441 48540: gap of 100 bp  
 \* 48541 49343: contig of 803 bp in length  
 \* 49344 49443: gap of 100 bp  
 \* 49444 50256: contig of 813 bp in length  
 \* 50257 50356: gap of 100 bp



Query Match		36 58; Score 365.2; DB 2; Length 77060;
Best Local Similarity		69.08; Pred. No. 1.le-769;
Matches 542; Conservative 0; Mismatches 209; Indels 35; Gaps 8;		
Qy	8	TCAGTTAGGAAGTAAGTCAAGAGATCTATTGTACTTGGTGACTACAGTAAATGTATT 67
Db	22530	TTAGACAGAGGAATAAGTCAAGAGATCTATTGACATGGTGACTACAGTAAATGTAT 22589
Qy	68	GTGTTCTTGA-----CTAATACAGTAGATTTCGAGTGTCTTCACACAAAACATGATG 121
Db	22590	CTGTTCTTGAACATGCTAGGACAGATAGATTGTAGTATTCTCACACAAAAA-AAGATG 22648
Qy	122	GATATGTCAGTAAATGCAAACTAGCTTGGTTAAACCATCCACATATGTGTGT 181
Db	22649	GGTTGTGAGGTAATGCAATAGCCAGTATCTTGGTTAAACCATCCATATGTGTGAT 22708
Qy	182	ATTTCAAAACAGTACCATAAATAGCAGACAAATTTTGTGTCAGTTACAAATCAAAAGTTT 241
Db	22709	ATTTCAACACAGTACCATAAAGTGTAGGCAATTTT-TATCAGTTACA-----AAAAAAATTT 22764
Qy	242	AAATGAGGACCTTAGGTGGTCTCTAATCCAAATCAAGTAAAGTGTCTCCATGAAAGGA 301
Db	22765	TAAATGAGGACCTTAGGAGAGAACCTTAATCCAATGTAACTGTGTCTCCACAAAGGAGA 22824
Qy	302	AATAGGATACAAATGTCCACACAGAGAGAAATGCCACATGAGGACACAAATGAGAATG 361
Db	22825	GATTAAGATACAAAGTCCACACAGAGAAATGCCACATGAGGACATAGAGAGAGAGGC 22884
Qy	362	GGCTACTTTACAGCCTTAGGAGAGGCGCTCCGAGAAAAACACACCCCTACCCACCTTGAT 421
Db	22885	GGCCACTTTACACCCAGGAGAGAGGCGCTCAGAGAAAAACACCTTGCCTCCACATCTTGAT 22944
Qy	422	GTGGACTTCACTCTGTAGACGAAGTCTCCACCCCTTTCATCAGGTGGAGGCTTTGA 481
Db	22945	CTTGGACTTCACT-----CTCCTTCATCAAAATGGAGGCTTCTGA 22983
Qy	482	TTCTGAATATTCTCAAAATCTGGAAGGTACAAAAGTCAAGACAGACAGACAGCTCAGG 541
Db	22984	TCCTGAATACTTTCAAAATCTAGAAAGTACAGTGGTGGAGGAGGACAGACAGAACTCANG 23043
Qy	542	GTGAAAGTTTAAAGAGAAATACATCTTTCCATTGCTGTGCTCTATCCCTTACACACACC 601
Db	23044	GTGAAAGTTTATAGAG-ATAACACCTCCCATTTGCCCTGTCTCTATCCCTCAACAC-CACC 23101
Qy	602	TATTCCAGTCTTTATTTGCTTTTGTGTTTTCGTTGCTGCTGGGTATAGTGTAGTTGTA 661

Db	23102	TGTGCCACGCTTTATTGTGTTTTCCTGATTCCTGCTGGGATAGTGTAAACTTNNNT 23161
Qy	662	TCTGTGT--TTACATACAGGATAAACAACAAAGGTAACAATAAATAAACAACAGACA 720
Db	23162	GATTNGTGTTCATAGATGATAACNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 23221
Qy	721	GCAAAACTCAACTAATAGTGTGTTGGGCATGTGACAGTGAAGACAGAGAGTCACATAA 780
Db	23222	NN 23281
Qy	781	AATGGA 786
Db	23282	NNNNGA 23287
RESULT 10		
AC005725/c		
LOCUS	AC005725	197915 bp DNA linear PRI 02-DEC-1998
DEFINITION	Homo sapiens chromosome 17, clone hRPK.680_C_21, complete sequence.	
ACCESSION	AC005725	
VERSION	AC005725.1	GI:3947428
KEYWORDS	HTG.	
SOURCE	Homo sapiens.	
ORGANISM	Homo sapiens	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 197915)	
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C., Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S., Collinsmore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K., Donelan,L., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S., Geraghty,K., Grant,G., Hagos,B., Heaford,A., Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L., Karatas,A., Lehoczy,J., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J., Nahr,R., Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A., Severy,P., Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C., Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D., Ye,W.J., Zhao,J. and Zody,M.	
TITLE	Direct Submission	
JOURNAL	Submitted (29-SEP-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	
REFERENCE	3 (bases 1 to 197915)	
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C., Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S., Collinsmore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K., Donelan,L., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S., Geraghty,K., Grant,G., Hagos,B., Heaford,A., Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L., Karatas,A., Lehoczy,J., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J., Nahr,R., Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A., Severy,P., Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C., Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D., Ye,W.J., Zhao,J. and Zody,M.	
TITLE	Direct Submission	
JOURNAL	Submitted (02-DEC-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	
COMMENT	On Dec 2, 1998 this sequence version replaced gi:393522. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html	

Only the last 197915 base pairs of this clone are being submitted.  
The remainder overlaps accession number AC005548 (WICGR project L361).

FEATURES		Location/Qualifiers
source		1..197915
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/chromosome="17"
		/map="17"
		/clone="hRPK_680_C-21"
		/clone_lib="RPC1-11 human BAC library"
repeat_region		complement(251..355)
repeat_region		/rpt_family="MIR"
repeat_region		687..969
repeat_region		/rpt_family="AluSc"
repeat_region		2640..2661
repeat_region		/rpt_family="AT_rich"
repeat_region		4789..5192
repeat_region		/rpt_family="MLT1B"
repeat_region		complement(5330..5795)
repeat_region		/rpt_family="MLT1H"
repeat_region		complement(7650..7676)
repeat_region		/rpt_family="TGAA)n"
repeat_region		complement(7722..8030)
repeat_region		/rpt_family="LTR16A"
repeat_region		complement(8051..8475)
repeat_region		/rpt_family="L2"
repeat_region		complement(8580..8676)
repeat_region		/rpt_family="MIR"
repeat_region		complement(8866..8935)
repeat_region		/rpt_family="MLT1I"
repeat_region		8936..8995
repeat_region		/rpt_family="TAGA)n"
repeat_region		complement(8998..9054)
repeat_region		/rpt_family="TAGA)n"
repeat_region		complement(9057..9381)
repeat_region		/rpt_family="MLT1I"
repeat_region		9391..9691
repeat_region		/rpt_family="AluSx"
repeat_region		9756..9824
repeat_region		/rpt_family="polypurine"
repeat_region		complement(9825..9847)
repeat_region		/rpt_family="MLT1J"
repeat_region		9848..9966
repeat_region		/rpt_family="GAA)n"
repeat_region		complement(9967..10060)
repeat_region		/rpt_family="MLT1J"
repeat_region		10947..11128
repeat_region		/rpt_family="L2"
repeat_region		11187..11570
repeat_region		/rpt_family="L2"
repeat_region		complement(12968..13015)
repeat_region		/rpt_family="(CGG)n"
repeat_region		13110..13159
repeat_region		/rpt_family="GC_rich"
repeat_region		complement(13396..13420)
repeat_region		/rpt_family="(CGG)n"
repeat_region		complement(13586..13654)
repeat_region		/rpt_family="(CGG)n"
repeat_region		13667..13758
repeat_region		/rpt_family="GC_rich"
repeat_region		13884..13927
repeat_region		/rpt_family="GC_rich"
repeat_region		13973..14023
repeat_region		/rpt_family="GC_rich"
repeat_region		18546..18848
repeat_region		/rpt_family="AluY"
repeat_region		complement(20108..20409)
repeat_region		/rpt_family="AluY"
repeat_region		complement(20687..20805)
repeat_region		/rpt_family="MIR"
repeat_region		21194..21495
repeat_region		/rpt_family="AluY"
repeat_region		complement(21620..21682)
repeat_region		/rpt_family="(CA)n"
repeat_region		complement(23260..23416)
repeat_region		/rpt_family="MIR"
repeat_region		26556..26614
repeat_region		/rpt_family="MIR"
repeat_region		27378..27513
repeat_region		/rpt_family="L2"
repeat_region		27644..27727
repeat_region		/rpt_family="purine-rich"
repeat_region		complement(28209..28257)
repeat_region		/rpt_family="L2"
repeat_region		complement(28225..28485)
repeat_region		/rpt_family="MIR"
repeat_region		28549..28847
repeat_region		/rpt_family="AluSx"
repeat_region		29715..30026
repeat_region		/rpt_family="AluSg"
repeat_region		30027..30068
repeat_region		/rpt_family="polypurine"
repeat_region		30328..30353
repeat_region		/rpt_family="AT_rich"
repeat_region		31266..31388
repeat_region		/rpt_family="(TGGA)n"
repeat_region		31474..31520
repeat_region		/rpt_family="purine-rich"
repeat_region		31537..31784
repeat_region		/rpt_family="MIR"
repeat_region		complement(33067..33188)
repeat_region		/rpt_family="purine-rich"
repeat_region		complement(33831..34379)
repeat_region		/rpt_family="L1MB6"
repeat_region		complement(34377..34527)
repeat_region		/rpt_family="L1"
repeat_region		complement(36437..36503)
repeat_region		/rpt_family="MIR"
repeat_region		complement(36546..36745)
repeat_region		/rpt_family="MIR"
repeat_region		37422..37595
repeat_region		/rpt_family="MER5B"
repeat_region		complement(37647..37836)
repeat_region		/rpt_family="MER5B"
repeat_region		38648..38746
repeat_region		/rpt_family="MER5A"
repeat_region		38818..39002
repeat_region		/rpt_family="MIR"
repeat_region		complement(39298..39373)
repeat_region		/rpt_family="MIR"
repeat_region		39598..39749
repeat_region		/rpt_family="MER5A"
repeat_region		complement(39764..39885)
repeat_region		/rpt_family="MER5B"
repeat_region		40689..41149
repeat_region		/rpt_family="L2"
repeat_region		complement(41364..41675)
repeat_region		/rpt_family="AluJb"
repeat_region		42215..42272
repeat_region		/rpt_family="MER3"
repeat_region		complement(42297..42607)
repeat_region		/rpt_family="AluSg"
repeat_region		complement(42851..43091)
repeat_region		/rpt_family="MER58A"
repeat_region		43312..43376
repeat_region		/rpt_family="MLT2D"
repeat_region		complement(43767..44095)
repeat_region		/rpt_family="AluSg"
repeat_region		44800..44932
repeat_region		/rpt_family="MIR"
repeat_region		complement(45393..45589)
repeat_region		/rpt_family="MIR"
repeat_region		complement(47813..48306)

Query Match

23.4%; Score 234.4; DB 9; Length 197915;

Best Local Similarity 73.3%; Pred. No. 7.5e-47;  
Matches 318; Conservative 0; Mismatches 106; Indels 10; Gaps 1;

QY 8 TCAGTTAGGAAGAAATGAAGTCGACAGATCTATTGTACTTGGTGACATACAGTTAATGCTATT 67  
DB 64448 TTGACAGGAGAAATGAAGTCGACAGATCTATTGTAATTATGGTTCT-----GTTTC 64399  
QY 68 GTGTTCTTGACTAATACAGTAGATTTCGAGTGTCTCACAACAAAAACATGATGGGTATG 127  
DB 64398 TTGAACGCTGCTAAGACAGTAGACTTTTCAGTGTCTCACAACAAAAAGATGATGGGTATG 64339  
QY 128 TGAGTAAATGCAATGCAAACTAGCTTGGGTAAACATTCACAAATATGTGTGTAATTTCA 187  
DB 64338 TGAGTAAACGATATGCTCATGCTGGTGGTGTAGCCATTCCACACGCGATGCATATTCA 64279  
QY 188 AAACAGTACCAATAATGACAGACAATTTGTGTCAGTTACAAATCAAAAAAGTTTTAAATG 247  
DB 64278 GAACAGTACCAATAATGAGAAAATTTTATCAGTTACAAATAAATAGTTTAAAT 64219  
QY 248 AGGACCTTAGGTGGTCTCTATCCAATCTAAGTGTCTCCATGAAAGAGAAATAAG 307  
DB 64218 GAGGACCTCGATGGCCCTAATCTAGTCAACTGATCTCCCTTAAAGAGAGGTAG 64159  
QY 308 GATCAAAATGTCACACAGAGAGAAATGGCCACATGAGGACACAATGAGATGTGGCTAC 367  
DB 64158 GATACAGATGTGCACACAGAGAGGGGTGACCCACATGAGGACAGAGGGGCGGCCAC 64099  
QY 368 TTACAAGCTAGGAGAGGGCTCCGAGAAACACACACCTACCCACACCTTGATTTGGA 427  
DB 64098 CTCAAGTCCGGGAGAGGGCTCAGAAGAACCCCTGCCACATCTTGTATTTGGA 64039  
QY 428 CTTCACTCTGTAGA 441  
DB 64038 TTCAACCTTCAGA 64025

RESULT 11  
AC025281  
LOCUS AC025281 193259 bp DNA linear PRI 30-MAR-2002  
DEFINITION Homo sapiens chromosome 16 clone RP11-356024, complete sequence.  
ACCESSION AC025281  
VERSION AC025281.7 GI:19848339  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
REFERENCE 1 (bases 1 to 193259)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
TITLE DOE Joint Genome Institute.  
JOURNAL Sequencing of Human Chromosome 16  
AUTHORS Unpublished  
REFERENCE 2 (bases 1 to 193259)  
DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (08-MAR-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
REFERENCE 3 (bases 1 to 193259)  
DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (30-MAR-2002) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
COMMENT On Mar 30, 2002 this sequence version replaced gi:13786380.  
Sequence Quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the  
GenBank flat file format but are available as part  
of this entry's ASN.1 file.  
-----  
Sequence Quality Assessment:

This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the  
GenBank flat file format but are available as part  
of this entry's ASN.1 file.  
-----

FEATURES  
source  
Location/Qualifiers  
1..193259  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="RP11-356024"  
BASE COUNT 53957 a 40929 c 42006 g 56367 t  
ORIGIN

Query Match 10.1%; Score 101.4; DB 9; Length 193259;  
Best Local Similarity 73.7%; Pred. No. 1.9e-14;  
Matches 157; Conservative 0; Mismatches 51; Indels 5; Gaps 2;

QY 219 TCAGTTACAAATCAAAAAAGTTTTAAATAGGACCTTAGGTGGTCTTAATCCCAATCTA 278  
DB 51145 TCAGTCTCAAAAAAGGTAAGTAAATAGTCTATTAGGTGGTCTTAATCCCAATAGA 51204  
QY 279 AGTGATGTCTCATCA-AGAGGAATAGGATACAATGTGCACACAGAGAAATGGC 337  
DB 51205 ACTGATGTGCTAAGAAGAGGAGCAAGGACACAGAT----ACACAGAGGAAAGAC 51260  
QY 338 CACATGAGGACACAAATGAGAAATGTGGCTACTTACAAAGCCTTAGGAGAGGCGCTCCGAGAA 397  
DB 51261 CTTGTGAAGACACAGAAGAGATGTCATTTACAAGCCAGGACGAGACGACTTCAGAAGA 51320  
QY 398 AACACACCTCCACACACCTTGATCTTGGACTT 430  
DB 51321 AATCAACCTGTCAACACCTTCATCTTGGACTT 51353

RESULT 12  
AC016716  
LOCUS AC016716 214269 bp DNA linear PRI 09-MAY-2001  
DEFINITION Homo sapiens BAC clone RP11-31213 from 2, complete sequence.  
ACCESSION AC016716  
VERSION AC016716.6 GI:11038573  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
REFERENCE 1 (bases 1 to 214269)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
TITLE Sulston, J.E. and Waterston, R.  
AUTHORS Toward a complete human genome sequence  
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
MEDLINE 99063792  
PUBMED 9847074  
REFERENCE 2 (bases 1 to 214269)  
Haakenson, B., Coblitz, B., Hawkins, M., Elliott, G., Abbott, A. and  
Boyer, E.  
TITLE The sequence of Homo sapiens BAC clone RP11-31213  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 214269)  
Waterston, R.H.  
AUTHORS Direct Submission  
TITLE Submitted (04-DEC-1999) Genome Sequencing Center, Washington  
JOURNAL University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
REFERENCE 4 (bases 1 to 214269)  
Waterston, R.H.  
AUTHORS Direct Submission  
TITLE Submitted (30-OCT-2000) Genome Sequencing Center, Washington  
JOURNAL University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

5 (bases 1 to 214269)  
Waterston, R.  
Direct Submission  
Submitted (09-MAY-2001) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Oct 30, 2000 this sequence version replaced gi:9910084.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
----- Summary Statistics  
Center project name: H\_NH0312I03  
-----

```
misc_feature 27556..28032
/note="similar to EST AA961265 (NID:g3127282) on96a03.s1"
misc_feature 27906..27955
/note="similar to EST AW593216 (NID:g7280474) hg11c01.x1"
repeat_region 28132..28330
/rpt_family="L1"
repeat_region 28366..28563
/rpt_family="L1"
misc_feature 28559..28952
/note="similar to EST AA082647 (NID:g1624764) zn24a06.rl"
repeat_region 28685..29040
/rpt_family="MaLR"
repeat_region 30275..30574
/rpt_family="Alu"
repeat_region 30687..31002
/rpt_family="Alu"
repeat_region 31148..31587
/rpt_family="L1"
repeat_region 31675..31748
/rpt_family="L1"
repeat_region 31749..32053
/rpt_family="Alu"

Query Match 9.8%; Score 97.6; DB 9; Length 214269;
Best Local Similarity 67.1%; Pred. No. 1.6e-13;
Matches 169; Conservative 0; Mismatches 79; Indels 4; Gaps 2;

QY 8 TCAGTTAGGAAGATAAGTGCAGAGATCT--ATTGACTTGGTGACTACAGTTAATGT 64
D 64534 TTATATAGGAGGATAAGTTCAGAGATCTAGCATCAACATGATGATATGTAATGT 64593

QY 65 ATT-GTGTCTTGGACTAATACAGTAGATTTTCAGTGTCTCACAACAAAAACATGATGG 123
D 64594 ATTCTTGAATAATGCTAAGAAGTAGATTTTAAGTGTTCGCATCACAAAAAATGAAAG 64653

QY 124 TATGTAGGTAATGCATATCAACTAGCTTGGGTTAACCAATTCACAAATATGTGTAT 183
D 64654 TATGTATATATGATATGATATGTAATAGCTTAATTTAGCCATTCACAAATGATATCATAT 64713

QY 184 TTCAAAACAGTACCATAATGCAGACAATTTTGTGTCAGTTACAAATCAAAAAAGTTTAA 243
D 64714 TTCAAAACATCATGTTTATATATATATTTCAAAACATCATGTTTATATATATATATTTCA 64773

QY 244 AATGAGGACCTT 255
D 64774 AAACATCATGTT 64785

RESULT 13
AC008632 112687 bp DNA linear PRI 27-JUN-2001
LOCUS Homo sapiens chromosome 5 clone CTB-164N12, complete sequence.
DEFINITION AC008632
ACCESSION AC008632
VERSION AC008632.6 GI:14572121
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 112687)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 112687)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 112687)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (27-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
```

```
COMMENT On Jun 27, 2001 this sequence version replaced gi:13699398.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www-shgc.stanford.edu
Quality: Phrap Quality >=40 99.8% of Sequence;
Estimated Total Number of Errors is 0.1.
STS Content:
WI-1102 G02844.
FEATURES             Location/Qualifiers
     source           1..112687
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTB-164N12"
BASE COUNT  32362 a 25636 c 24363 g 30326 t
ORIGIN
Query Match 9.6%; Score 96.2; DB 9; Length 112687;
Best Local Similarity 76.2%; Pred. No. 3.6e-13;
Matches 147; Conservative 0; Mismatches 38; Indels 8; Gaps 2;

QY 8 TCAGTTAGGAAGATAAGTGCAGAGATCTATTGTGCTGCTACAGTTAATGTATT 67
D 103664 TTACACAGGAGGATAAATGTTCAAGAGATATACTGTCAACATGATGATGTAATGATT 103723

QY 68 GTGTTCTTGG-----ACTAATACAGTAGATTTTCGAGTGTCTCTCAAC--AAAAACATGA 119
D 103724 GTATCTTTAAATCACTAAGGGGTAGATTTTAAAGTGTCTTCCACCACCAAAAAATGG 103783

QY 120 TGGGTATGTGAGGTAATGCATATCAAACTAGCTTGGGTTAACCAATTCACAAATGTGT 179
D 103784 TAAGTATGTGAGGTAATGCATATGTTAATAGCTCAATTTAGCCATTCACCAATGTCTAT 103843

QY 180 GTATTTCAAAACA 192
D 103844 GTATTTCAAAACA 103856

RESULT 14
AC022259 195470 bp DNA linear HTG 27-JAN-2000
LOCUS Homo sapiens chromosome 2 clone RP11-335E2 map 2, *** SEQUENCING IN
DEFINITION AC022259
ACCESSION AC022259
VERSION AC022259.1 GI:6778490
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 195470)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 2, clone RP11-335E2
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 195470)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bada,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepeil,K., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArelano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferrelira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lechoky,J., Levine,R., Liu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheters,R., Meldrum,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
```

```

TITLE      Direct Submission
JOURNAL    Submitted (27-JAN-2000) Whitehead Institute/MIT Center for Genome
COMMENT    Research, 320 Charles Street, Cambridge, MA 02141, USA
           All repeats were identified using RepeatMasker:
           Smit, A.F.A. & Green, P. (1996-1997)
           http://ftp.genome.washington.edu/RM/RepeatMasker.html
           ----- Genome Center
           Center: Whitehead Institute/ MIT Center for Genome Research
           Center code: WB9R
           Web site: http://www-seq.wi.mit.edu
           Contact: sequence_submissions@genome.wi.mit.edu
           ----- Project Information
           Center project name: L5217
           Center clone name: 335_E_2
           -----
           * NOTE: This is a 'working draft' sequence. It currently
           * consists of 58 contigs. The true order of the pieces
           * is not known and their order in this sequence record is
           * arbitrary. Gaps between the contigs are represented as
           * runs of N, but the exact sizes of the gaps are unknown.
           * This record will be updated with the finished sequence
           * as soon as it is available and the accession number will
           * be preserved.
           *
           * 1 1169: contig of 1169 bp in length
           *      gap of unknown length
           * 1170 2744: contig of 1575 bp in length
           *      gap of unknown length
           * 2745 3891: contig of 1147 bp in length
           *      gap of unknown length
           * 3892 4958: contig of 1067 bp in length
           *      gap of unknown length
           * 4959 6042: contig of 1084 bp in length
           *      gap of unknown length
           * 6043 7112: contig of 1070 bp in length
           *      gap of unknown length
           * 7113 8242: contig of 1130 bp in length
           *      gap of unknown length
           * 8243 9440: contig of 1198 bp in length
           *      gap of unknown length
           * 9441 10503: contig of 1063 bp in length
           *      gap of unknown length
           * 10504 11766: contig of 1263 bp in length
           *      gap of unknown length
           * 11767 12798: contig of 1032 bp in length
           *      gap of unknown length
           * 12799 14299: contig of 1501 bp in length
           *      gap of unknown length
           * 14300 15314: contig of 1015 bp in length
           *      gap of unknown length
           * 15315 16978: contig of 1664 bp in length
           *      gap of unknown length
           * 16979 18049: contig of 1071 bp in length
           *      gap of unknown length
           * 18050 19597: contig of 1548 bp in length
           *      gap of unknown length
           * 19598 21003: contig of 1406 bp in length
           *      gap of unknown length
           * 21004 22126: contig of 1123 bp in length
           *      gap of unknown length
           * 22127 23502: contig of 1376 bp in length
           *      gap of unknown length
           * 23503 24450: contig of 948 bp in length
           *      gap of unknown length
           * 24451 25766: contig of 1316 bp in length
           *      gap of unknown length
           * 25767 27235: contig of 1469 bp in length
           *      gap of unknown length
           * 27236 29226: contig of 1991 bp in length
           *      gap of unknown length
           * 29227 30961: contig of 1735 bp in length
           *      gap of unknown length
           * 30962 32608: contig of 1647 bp in length
           *      gap of unknown length
           *
           * 32609 34331: contig of 1723 bp in length
           *      gap of unknown length
           * 34332 36533: contig of 2202 bp in length
           *      gap of unknown length
           * 36534 39299: contig of 2766 bp in length
           *      gap of unknown length
           * 39300 41451: contig of 2152 bp in length
           *      gap of unknown length
           * 41452 43295: contig of 1844 bp in length
           *      gap of unknown length
           * 43296 45678: contig of 2383 bp in length
           *      gap of unknown length
           * 45679 48551: contig of 2873 bp in length
           *      gap of unknown length
           * 48552 51186: contig of 2635 bp in length
           *      gap of unknown length
           * 51187 53207: contig of 2021 bp in length
           *      gap of unknown length
           * 53208 55697: contig of 2490 bp in length
           *      gap of unknown length
           * 55698 59328: contig of 2631 bp in length
           *      gap of unknown length
           * 59329 59424: contig of 1096 bp in length
           *      gap of unknown length
           * 59425 62209: contig of 2785 bp in length
           *      gap of unknown length
           * 62210 64795: contig of 2586 bp in length
           *      gap of unknown length
           * 64796 67972: contig of 3177 bp in length
           *      gap of unknown length
           * 67973 71707: contig of 3735 bp in length
           *      gap of unknown length
           * 71708 74928: contig of 3221 bp in length
           *      gap of unknown length
           * 74929 78667: contig of 3739 bp in length
           *      gap of unknown length
           * 78668 82959: contig of 4292 bp in length
           *      gap of unknown length
           * 82960 88848: contig of 5889 bp in length
           *      gap of unknown length
           * 88849 93807: contig of 4959 bp in length
           *      gap of unknown length
           * 93808 98630: contig of 4823 bp in length
           *      gap of unknown length
           * 98631 106830: contig of 8200 bp in length
           *      gap of unknown length
           * 106831 114298: contig of 7468 bp in length
           *      gap of unknown length
           * 114299 121722: contig of 7424 bp in length
           *      gap of unknown length
           * 121723 128019: contig of 6297 bp in length
           *      gap of unknown length
           * 128020 136332: contig of 8313 bp in length
           *      gap of unknown length
           * 136333 144730: contig of 8398 bp in length
           *      gap of unknown length
           * 144731 153111: contig of 8381 bp in length
           *      gap of unknown length
           * 153112 162589: contig of 9478 bp in length
           *      gap of unknown length
           * 162590 171983: contig of 9394 bp in length
           *      gap of unknown length
           * 171984 181733: contig of 9750 bp in length
           *      gap of unknown length
           * 181734 195470: contig of 13737 bp in length.
           *
           * Location/Qualifiers
           *   1..195470
           *     /organism="Homo sapiens"
           *     /db_xref="taxon:9606"
           *     /chromosome="2"
           *     /map="2"
           *     /clone="RP11-335E2"
           *     /clone_lib="RPC1-11 Human Male BAC"
           *
           FEATURES
           source

```

```

BASE COUNT  54946 a 42867 c 42439 g 55210 t      8 others
ORIGIN
Query Match          9.6%; Score 96.2; DB 2; Length 195470;
Best Local Similarity 76.2%; Pred. No. 3.5e-13;
Matches 147; Conservative 0; Mismatches 36; Indels 8; Gaps 2;

QY  8 TCAGTTAGGAGAATAAGTGCAGAGATCTATTGTGACTACAGTTTAATGTATT 67
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 110474 TTAGACAGAGGAATAAGTTCAAGAGATATACTGTACAACATGATGATTAATGTATT 110533

QY  68 GGTCTTCTG-----ACTAATACACATAGATTTTCGAGTGTCTCACAAC--AAAAACATGA 119
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 110534 GTATTCTTTAAATCACAAGGGGTAGATTTTAAGTGTCTTCACACCAAAAAAATGG 110593

QY  120 TGGGTATGTGAGGTAAATGCATATGCAAACTAGCTTGGGTTAACCAATTCACAAATATGTGT 179
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 110594 TAAGTATGTGAGGTAAATGCATATGTTAATGCTCAATTTAGCCATTCACAAATGCTAT 110653

QY  180 GTATTTCAAAACA 192
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 110654 GTATTTCAAAACA 110666

RESULT 15
AC093514      147330 bp      DNA      linear      PRI 04-JAN-2002
LOCUS      Homo sapiens chromosome 16 clone CTD-2517022, complete sequence.
DEFINITION      AC093514
ACCESSION      AC093514
VERSION      AC093514.2 GI:18057084
KEYWORDS      HTG.
SOURCE      Homo sapiens.
ORGANISM      Homo sapiens
               Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 147330)
               DOE Joint Genome Institute.
               Sequencing of Human Chromosome 16
               Unpublished
REFERENCE      2 (bases 1 to 147330)
               DOE Joint Genome Institute.
               Direct Submission
               Submitted (31-AUG-2001) Production Sequencing Facility, DOE Joint
               Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
               3 (bases 1 to 147330)
               DOE Joint Genome Institute.
               Direct Submission
               Submitted (04-JAN-2002) Production Sequencing Facility, DOE Joint
               Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
               On Jan 4, 2002 this sequence version replaced gi:15383803.
COMMENT      Sequence Quality Assessment:
               This entry has been annotated with sequence quality
               estimates computed by the Phrap assembly program.
               All manually edited bases have been reduced to quality zero.
               Quality levels above 40 are expected to have less than
               1 error in 10,000 bp.
               Base-by-base quality values are not generally visible from the
               GenBank flat file format but are available as part
               of this entry's ASN.1 file.
               -----
               Sequence Quality Assessment:
               This entry has been annotated with sequence quality
               estimates computed by the Phrap assembly program.
               All manually edited bases have been reduced to quality zero.
               Quality levels above 40 are expected to have less than
               1 error in 10,000 bp.
               Base-by-base quality values are not generally visible from the
               GenBank flat file format but are available as part
               of this entry's ASN.1 file.
               -----
FEATURES             Location/Qualifiers
     source            1..147330
                       /organism="Homo sapiens"

```

```

/db_xref="taxon:9606"
/chromosome="16"
/clone="CTD-2517022"
BASE COUNT  41780 a 30102 c 30898 g 44550 t
ORIGIN
Query Match          9.5%; Score 95.2; DB 9; Length 147330;
Best Local Similarity 71.9%; Pred. No. 6.3e-13;
Matches 138; Conservative 0; Mismatches 53; Indels 1; Gaps 1;

QY  240 TTAATAATGAGGACCTTAGGGTGGTCCCTAATCAATCTAAGTGATGTCTCCATGAAGAG 299
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 119670 TTAATAATGAGGCCCTTAGGGTGGGCCCTAATCAATCTGCCTTTATGAGAGAG 119729

QY  300 GAAATAAGGATACAAATGTGCACACAGAGAGAAATGGCCACATGAGGACACATGAGAAT 359
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 119730 GAAAGACACACCAGAGATGTGCACTCACAGAGGGGGCCATGTGAGGACACAGCAAGAG 119789

QY  360 GTGGCTACTTTACAAGCCCTAGGAGAGAGGCCCTCCGAGAAAAACACACCTTACCACACCTTG 419
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 119790 GTGGCCATCCGCAAGCAAGGAGAGAGG-CTTCAAGAGAATCTAAACCTGTGTCACCTTG 119848

QY  420 ATGTTGGACTTC 431
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 119849 ATCTGGGACTTC 119860

Search completed: December 24, 2002, 21:06:13
Job time : 3113.19 secs

```

**THIS PAGE BLANK (uspto)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 24, 2002, 16:24:42 ; Search time 1812.02 Seconds  
(without alignments)  
16077.036 Million cell updates/sec

Title: US-09-708-724A-3\_COPY\_70000\_71000  
Perfect score: 1001  
Sequence: 1 ggagatgataaacggtgtg.....ccaattcaggagtctatgtg 1001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_rst.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_in.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pi.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1001	100.0	191768	9	AC092357	AC092357 Homo sapi
2	992.4	99.8	151049	9	AC018558	AC018558 Homo sapi
c 3	836.8	83.6	191014	9	AC113268	AC113268 Papio cyn
4	835	83.4	185664	9	AC093249	AC093249 Homo sapi
5	726.8	72.6	153875	9	AC003682	AC003682 Homo sapi
6	649.2	64.9	45383	9	AC010642	AC010642 Homo sapi
7	649.2	64.9	160852	9	AC020915	AC020915 Homo sapi
8	649.2	64.9	214042	2	AC023149	AC023149 Homo sapi
9	627.4	62.7	158802	2	AC015589	AC015589 Homo sapi
10	625.8	62.5	196686	9	AC010328	AC010328 Homo sapi
c 11	615	61.4	124102	9	AC021148	AC021148 Homo sapi
12	615	61.4	158302	2	AC069481	AC069481 Homo sapi
c 13	602	60.1	4473	9	AF017337	AF017337 Homo sapi
c 14	590.2	59.0	153937	9	AL451075	AL451075 Human DNA
c 15	589	58.8	133704	9	AC008567	AC008567 Homo sapi
c 16	584.6	58.4	131826	9	AC011460	AC011460 Homo sapi
c 17	583.6	58.3	106648	2	AC004588	AC004588 Homo sapi
c 18	583.6	58.3	159849	2	AC021454	AC021454 Homo sapi
c 19	583.6	58.3	163915	2	AC087451	AC087451 Homo sapi
c 20	583.6	58.3	166867	9	AP003733	AP003733 Homo sapi
c 21	579.6	57.9	184169	2	AC026349	AC026349 Homo sapi
c 22	579.6	57.9	219553	2	HS312687	HS312687 Homo sapi
c 23	574.8	57.4	160169	2	AC051664	AC051664 Homo sapi
c 24	571.6	57.1	137830	9	AC003029	AC003029 Homo sapi
c 25	562	56.1	196431	2	CNS01DU5	AL133162 Homo sapi
c 26	557.2	55.7	115732	9	AC104566	AC104566 Homo sapi
c 27	557.2	55.7	158535	9	AC093368	AC093368 Homo sapi
c 28	557	55.6	124047	9	AL713966	AL713966 Human DNA
c 29	556.4	55.6	179581	2	AC009397	AC009397 Homo sapi
c 30	555.2	55.5	80611	9	HS323P24	AL022157 Human DNA
c 31	553.2	55.3	35715	9	AC020951	AC020951 Homo sapi
c 32	546.4	54.6	170470	9	AC092364	AC092364 Homo sapi
c 33	546.4	54.6	182154	2	AC041049	AC041049 Homo sapi
c 34	544.8	54.4	129794	9	AL353812	AL353812 Human DNA
c 35	544.8	54.4	132229	9	HS447N6	282211 Human DNA S
c 36	544.8	54.4	171970	2	AL590454	AL590454 Homo sapi
c 37	544	54.3	179272	2	AC069149	AC069149 Homo sapi
c 38	538.8	53.8	104082	9	AC011503	AC011503 Homo sapi
c 39	526.8	52.6	38757	9	AC008976	AC008976 Homo sapi
c 40	524.4	52.4	149794	2	AC007273	AC007273 Homo sapi
c 41	524.4	52.4	161879	9	AC017005	AC017005 Homo sapi
c 42	522.8	52.2	120951	9	AC073649	AC073649 Homo sapi
c 43	520.6	52.0	164295	2	AC025958	AC025958 Homo sapi
c 44	520.6	52.0	196826	9	AC026775	AC026775 Homo sapi
c 45	520.4	52.0	187280	9	AC092674	AC092674 Homo sapi

ALIGNMENTS

RESULT 1

AC092357

LOCUS

DEFINITION

AC092357 AC023764

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

REFERENCE

AUTHORS

TITLE

JOURNAL

AC092357 191768 bp DNA linear PRI 15-AUG-2001  
Homo sapiens chromosome 16 clone RP11-332P24, complete sequence.

AC092357 AC023764

AC092357.2 GI:15187270

HTG.

Homo sapiens.

Homo sapiens

Eukaryota: Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 191768)

DOE Joint Genome Institute.

Sequencing of Human Chromosome 16

Unpublished



Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

-----  
Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.

All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

-----  
FEATURES

source

Location/Qualifiers

1. .151049

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="16"

/clone="RP11-80F22"

BASE COUNT 46827 a 30783 c 30749 g 42690 t

ORIGIN

Query Match 99.88; Score 999.4; DB 9; Length 151049;  
Best Local Similarity 99.9%; Pred. No. 1.1e-299;  
Matches 1000; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGAGATGATAAACCGTGTAGTGCCTCAAGTTGTGTGGCACCATTGGAATGGGAGACTG 60  
Db 137896 GGAGATGATAAACCGTGTAGTGCCTCAAGTTGTGTGGCACCATTGGAATGGGAGACTG 137955

QY 61 GAGGATACATGGATGCCCAACTACAGGCCCGCCTCCAGTATGAGCCATGAGCCAGTT 120  
Db 137956 GAGGATACATGGATGCCCAACTACAGGCCCGCCTCCAGTATGAGCCATGAGCCAGTT 138015

QY 121 GAATCTGAATGTGAATGGAATGAAGCAGCAGAGAGTACACAGTCAACCCCTCAT 180  
Db 138016 GAATCTGAATGTGAATGGAATGAAGCAGCAGAGTACACAGTCAACCCCTCAT 138075

QY 181 AACATGGGTGATCAAGAAACACACACAGAGCTGAGAACTGGTGTAGTCCAGGG 240  
Db 138076 AACATGGGTGATCAAGAAACACACACAGAGCTGAGAACTGGTGTAGTCCAGGG 138135

QY 241 TCAGGCAAAACCCCTGACTCCATGTTTATGGCCATGCTAGCTGTAATATCTGTGCAGT 300  
Db 138136 TCAGGCAAAACCCCTGACTCCATGTTTATGGCCATGCTAGCTGTAATATCTGTGCAGT 138195

QY 301 ATGATTTTCTGTGAGAGCAAAACATATTTGGGCATATTTTCTTAACCCACCGGTAGT 360  
Db 138196 ATGATTTTCTGTGAGAGCAAAACATATTTGGGCATATTTTCTTAACCCACCGGTAGT 138255

QY 361 GTGATCATCTCTGAAGCAGACATCTCTCTGAGATATATCATATGATCAAGGAGCATCAGTA 420  
Db 138256 GTGATCATCTCTGAAGCAGACATCTCTCTGAGATATATCATATGATCAAGGAGCATCAGTA 138315

QY 421 CCAGGACCTCTAACTCCCGCTGACACAGAGCAATTAGACTCTCTATAACAATGGTATCAAT 480  
Db 138316 CCAGGACCTCTAACTCCCGCTGACACAGAGCAATTAGACTCTCTATAACAATGGTATCAAT 138375

QY 481 TATACCACTCATTTGGAGGAGCTTCTTTTATGTGTACCCAGGATACATGCTCAACTGC 540  
Db 138376 TATACCACTCATTTGGAGGAGCTTCTTTTATGTGTACCCAGGATACATGCTCAACTGC 138435

QY 541 AGTTGCCCTTGAGTTGATCCCAAGCATGGTTGAGTTACCATAAAATAATATCTACCTA 600  
Db 138436 AGTTGCCCTTGAGTTGATCCCAAGCATGGTTGAGTTACCATAAAATAATATCTACCTA 138495

QY 601 TTAGACCTTAGCTTTATTAATATATCTGTGTAGTTACTATCACTCTCTGCCCCCATCAC 660  
Db 138496 TTAGACCTTAGCTTTATTAATATATCTGTGTAGTTACTATCACTCTCTGCCCCCATCAC 138555

QY 661 CCAAAATGTACTGATTTACAGAAATGGGCTCCCTTTGATATTTCTACCCCTCTCTTGG 720  
|||||

Db 138556 CCAAAATGTACTGATTTACAGAAATGGGCTCCCTTTGATAAATCTTCACCCCTCTCTGG 138615

QY 721 GCCCAGCTGTCTTGGCCCTTAGCTAGACAATAGTCCATGTTAATGGGAGACATATTTCAC 780  
|||||

Db 138616 GCCCAGCTGTCTTGGCCCTTAGCTAGACAATAGTCCATGTTAATGGGAGACATATTTCAC 138675

QY 781 TGGGTCCCTCTGTGTCATTAAAGATGGGAGAGATGAGAATCAGACACATGGCATAAACTT 840  
|||||

Db 138676 TGGGTCCCTCTGTGTCATTAAAGATGGGAGAGATGAGAATCAGACACATGGCATAAACTT 138735

QY 841 CACTGGCACTGGTGGCGAACTTTAAACATCTCTTCACTTCAACACACTGGGATCAATCC 900  
|||||

Db 138736 CACTGGCACTGGTGGCGAACTTTAAACATCTCTTCACTTCAACACACTGGGATCAATCC 138795

QY 901 CAATCTGCCATGCAACTTGTTCATGCAATGCAAGCGCTTTAGCCACCTTCCTCAATGG 960  
|||||

Db 138796 CAATCTGCCATGCAACTTGTTCATGCAATGCAAGCGCTTTAGCCACCTTCCTCAATGG 138855

QY 961 CATTATCAAGGAAAGAGAGGTCCCAATTCAGGAGTCTATGTG 1001  
|||||

Db 138856 CATTATCAAGGAAAGAGAGGTCCCAATTCAGGAGTCTATGTG 138896

## RESULT 3

AC113268/c

LOCUS

AC113268

Papio cynocephalus anubis clone rp41-22m16, complete sequence.

AC113268

VERSION

AC113268.8

GI:22095124

HTG.

KEYWORDS

HTG.

SOURCE

olive baboon.

ORGANISM

Papio cynocephalus anubis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;

Cercopithecinae; Papio.

REFERENCE

1 (bases 1 to 191014)

Zhou,L., Fu,Y., Wu,J., Eichler,E. and Roe,B.A.

Papio anubis BAC Clone rp41-22m16

Unpublished

REFERENCE

2 (bases 1 to 191014)

Zhou,L., Fu,Y., Wu,J., Eichler,E. and Roe,B.A.

Direct Submission

Submitted (06-AUG-2002) Department Of Chemistry And Biochemistry,

The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,

OK 73019, USA

REFERENCE

3 (bases 1 to 191014)

Zhou,L., Fu,Y., Wu,J., Eichler,E. and Roe,B.A.

Direct Submission

Submitted (06-AUG-2002) Department Of Chemistry And Biochemistry,

The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,

OK 73019, USA

REFERENCE

4 (bases 1 to 191014)

Zhou,L., Fu,Y., Wu,J., Eichler,E. and Roe,B.A.

Direct Submission

Submitted (14-AUG-2002) Department Of Chemistry And Biochemistry,

The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,

OK 73019, USA

COMMENT

On Aug 4, 2002 this sequence version replaced gi:21104942.

----- Genome Center

Center: Department Of Chemistry And Biochemistry

The University Of Oklahoma

Center code:UOKNOR

-----

Location/Qualifiers

1. .191014

/organism="Papio cynocephalus anubis"

/sub\_species="anubis"

/db\_xref="taxon:9555"

/clone="rp41-22m16"

/clone\_lib="RPCI - 41 Male (Olive) Baboon BAC Library"

54882 a 40134 c 39560 g 56438 t

BASE COUNT

ORIGIN

```
Query Match      83.6%; Score 836.8; DB 9; Length 191014;
Best Local Similarity 91.5%; Pred. No. 5.6e-249;
Matches 922; Conservative 0; Mismatches 77; Indels 9; Gaps 3;

QY 1 GGAGATGATAAACCGCTGTGAGTGCCTCAAGTTCTGTGCGACCATGGAATGGGAGACTG 60
Db 109942 GGAGATGATAAACCATGTGAGTGCCTCAAGGTGTGTGTGACCATGGAATGGGAGACTG 109883

QY 61 GAGGATACATGGAATCCCACTACAGGCCAGCTCCTCCAGTATGAGCCATGAGCCAGTT 120
Db 109882 GAGGATACATGGAATCCCACTACAGGCTTGGCTCCTCCAGTATGAGCCATGAGCCAGTT 109823

QY 121 GAATCTGAATGTGAAGATGGAATGAAGACCGACGAGATCACTAGCGTCAACCCCTCAT 180
Db 109822 GAATCTGAATGTGAAGATGGAATGAAGACTGACGAGGACATTTGACATCAACCCCTCAT 109763

QY 181 AACATGGGGTCAGATCAAGAAACACACACAGAGCTGAGAACTGGTGTAGTGCCAGGG 240
Db 109762 AACGTGGGTACGGTTCAAGAAACACACAGAGCTGAGAACTGGTGGAGTGCCAGGG 109703

QY 241 TCAGGCAAAAACCCCTGACTCCTATGTTATGGCCATGCTAGCTGTAATATCCTGTGCAGT 300
Db 109702 TCAGGCAAAAACCCCTGACTCCTATGTTATGGCCATGCTAGCTGTAATATCCTGTGCAGT 109643

QY 301 ATGATTTTCTGTGCAGAGCAAAACATATTTGGGCATATTTTCTCAACCCACCGGTAGT 360
Db 109642 ATGATTTTCTGTGCAGAGCAAAACATATTTAGTGTATGTTCTCCACCCACCGGTAGT 109583

QY 361 GTGATCATACTCTGAAGCAGCACTCCTCTGAGATATATCATGATCAAGGAGCATCAGTA 420
Db 109582 GTGACCATACTCTGAAGTGCACACTCCTCTGAGGTATATCATGATCAAGGAGTATAGTA 109523

QY 421 CCAGGACCTTAACTCCCTGACACAGCAATAGACTCTCTCAFAACAATGGTATCAAT 480
Db 109522 CCAGGACCTTAACTCCCTGACACAGCAATAGACTCTCTCAFAACAATGGTATCAAT 109463

QY 481 TATACCATCCATTTGAGGAGACTCCTTTATGTGTACCCAGGATACATTTGCTCAACTGC 540
Db 109462 TATATCACTCCATT--GAGGAGACTCCTTTATGTGTATGATCAACAGGATGATCAACTGC 109404

QY 541 AGTTCCTTGCAGTTGATCCCAAGCATGTTGAGTTACCA-----TAAATAAATTAT 593
Db 109403 AGTTCCTTGCAGTTGATCCCAAGCATGTTGAGTTACCAAAAAAAAAAAAAAATTAT 109344

QY 594 GTACCTATTAGACCTTAGCTTTATTAATATTACTTGTGTAGTACTTAATCACTCTCTGGCC 653
Db 109343 GTACCTATTAGACCTTAGCTTTATTAATATTATTGTTGTAGTACTTAATCACTCTCTGGCC 109284

QY 654 CCATCACCACCAATTTGATGATTTATACAGATGGGCTCCTTTGATATTAATTCACCCGCC 713
Db 109283 CCATCACCACCAATTTGCTGATTTATACAGATGGGCTCCTTTGATATTAATTCAC--CCGCC 109225

QY 714 TCCTTTGGGCCCACTCTCTTGGCCCTTAGCTAGACAATAGTCCATGTTAATGGGAGACAT 773
Db 109224 TCCTTTGGGCCCACTCTCTTGGGCCCACTAGCTAGACAATAGTCCATGTTAATGGAAGACAT 109165

QY 774 TATTGACTGGGTCCTCTGTGTGTCATTAAGATGGGAGATGAGATGAGATCAACACATGGCA 833
Db 109164 TATTGACTGGGTCCTCTGTGTGTCATTTAGATAGGAGATGAGATGAGATCAACCTCATGGCA 109105

QY 834 TAAACTTCACCTGGCACTGGTGGCGAACTTTTAACTATCTTTCACATTCACACACTGGGAT 893
Db 109104 TAAACTTCACCTGGTGGTGGAACTTTTAACTATCTTTCGCTTACACACACTGGGAT 109045

QY 894 TCAATCCCAATCTGCCATCACTTGTGGCATGGAAGGCTTTAGCCACCTTTGCC 953
Db 109044 TCAATCCCAATCTGCCATCACTTGTGGCATGGAAGGCTTTAGCCACCTTTGCC 109885

QY 954 TCAATGGCATTTAAGGAAAGAGAGGTCCTCAATTCAGGAGTCTATGCT 1001
Db 109884 TCAATGGCATTTAAGGAAAGAGAGGTCCTCAATTCAGGAGTCTATG 108937
```

RESULT 4  
AC093249 185664 bp DNA linear PRI 03-FEB-2002  
LOCUS Homo sapiens chromosome 16 clone RP11-146F11, complete sequence.  
DEFINITION AC093249  
ACCESSION AC093249  
VERSION AC093249.3 GI:18481999  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 185664)  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Human Chromosome 16  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 185664)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL DOE Joint Genome Institute.  
REFERENCE 3 (bases 1 to 185664)  
AUTHORS Submitted (16-AUG-2001) Production Sequencing Facility, DOE Joint  
TITLE Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
JOURNAL  
REFERENCE 4 (bases 1 to 185664)  
AUTHORS Submitted (03-FEB-2002) Production Sequencing Facility, DOE Joint  
TITLE Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
JOURNAL  
COMMENT Sequence Quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the  
GenBank flat file format but are available as part  
of this entry's ASN.1 file.  
-----  
Sequence Quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the  
GenBank flat file format but are available as part  
of this entry's ASN.1 file.  
-----

FEATURES  
source  
1. 185664  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="RP11-146F11"

BASE COUNT 45014 a 47979 c 46282 g 46389 t

ORIGIN  
Query Match 83.4%; Score 835; DB 9; Length 185664;  
Best Local Similarity 90.4%; Pred. No. 2e-248;  
Matches 892; Conservative 0; Mismatches 95; Indels 0; Gaps 0;  
QY 1 GGAGATGATAAACCGTGTGAGTGCCTCAAGTTCTGTGCGACCATGGAATGGGAGACTG 60  
Db 88657 GGAGATGACAAACCGTGCCTGTAGGTGTGTGCGACCATGGAATGGGAGACTG 88716  
QY 61 GAGGATACATGGAATCCCACTACAGGCCAGCTCCTCCAGTATGAGCCATGAGCCAGTT 120  
Db 88717 GAGGATCTGTGGATCCCACTACCGTCCGGTTCCTCCCATGATGAGCCATGAGCCAGTT 88776  
QY 121 GAATCTGAATGTGAAGATGGAATGAAGACCGACGAGATCACTAGCGTCAACCCCTCAT 180  
Db 88777 GAATCTGAATGCGAAGACAGAATGGGGGCCAACCCGCGTCAACATGACATCAACCCCTCAT 88836  
QY 181 RACATGGGGTCAGATCAAGAAACACACACAGAGCTGAGAACTGGTGTAGTGCCAGGG 240

Db	88837	ANATGGGGACAGATCAAGAAAAACACACAGAGAAAGTTGAAAAACTGCTGGAGCGCCAGGG	88896
QY	241	TCAGGCAAAACCCCTGACTCCATGCTTTATGGCCATGCTAGCTGTAATATCTGTGCACT	300
Db	88897	TCAGGCAAAACCCAGACTCCTATGTTATGGCCATGCTAGCTGTAATATCTGTGAGT	88956
QY	301	ATGATTTTCTGTGCAGAGCAAAACATATTTGGGCATATTTTCCTAACCCACCGGTAGT	360
Db	88957	ATGTTTTCCTGTGCAGAGGCAAAACATATTTGGGCATATTTTCCTAACCCACTGGTAGT	89016
QY	361	GTGATCATACTCTGAAGCAGCACTCCTCTGAGATATATCATGATCAAGGACATCAGTA	420
Db	89017	GTGACATACTCTGAACGACCACTCCTCTGAGATATATCATGATCAAGGACATGAGCA	89076
QY	421	CCAGGACCTCTAACCTCCCTGCACAGAGCAATTAGACTCTCATAAACAATGGTATCAAT	480
Db	89077	CCAGGACCTCTAACCTCCCTGCACAGAGCAATTAGACTCTCATAGACAATGGTATCAAT	89136
QY	481	TATACACTCCATTTGGAGGACTTCCTTTATGTGTCACCCAGGATACATGTCCTCACTGC	540
Db	89137	TATACACTCCATTTGGAGGAACTTCCTTTATGCGTCACCCAGGATACCTCACTCACTGC	89196
QY	541	AGTTGCTTGGCAGTTTGATCCCAAGCATGGTTGAGTTACCATATAAAAAAATTATGTACCTA	600
Db	89197	AGTTGCTTGGCAATTCATCCCAAGCATGGCTGAGTTACCATGGAATATTTATGTACCTA	89256
QY	601	TTAGACCTTAGCTTATTAATATTAATCTGTGTAGTTACTAATCACTCTCTGGCCCATCAC	660
Db	89257	TTAGGCTTAGCTTATTAATATTAATCTGTGTAGTTACTAATCACTCTCTGGCCCATCGC	89316
QY	661	CCAAATTTGACTGATATACAGAATGGGCTCCCTTTGATATTTCTACCCCCCTCTCTGG	720
Db	89317	CCAAATTTGACTGATATATACAGAATGGGCTCCCTTTGATATTTCTACCCCCCTCTCTGG	89376
QY	721	GCCCACTGTTCGCCCTTACTAGACATAATAGTCCATCTTAATGGGACATATTATGAC	780
Db	89377	GCCCACTGTTCGCCCTTACTAGACATAATAGTCCATCTTAATGGGACATATTATGAC	89436
QY	781	TGGGTCCCTGTGTCATTAAAGATGGGAGATGAGATCAGACACATGCGCATAAACTT	840
Db	89437	TGGGTCCCTGTGTCATTAAAGATGGGAGATGAGATCAGACACATGCGCATAAACTT	89496
QY	841	CACTGGCACTGGTGGCAAACTTTAACTCTCTTCACTTCAACACACATGGGATCAATCC	900
Db	89497	CACTGGCACTGGTGGCAAACTTTAACTCTCTTCACTTCAACACACATGGGATCAATCC	89556
QY	901	CAATCTGCATGCACTTGTGGCATGGAAGGCTTATAGCCACCTTTGCTCTCAATGG	960
Db	89557	CAATCTGCATGCACTTGTGGCATGGAAGGCTTATAGCCACCTTTGCTCTCAATGG	89616
QY	961	CATTATCAAGGAAAGAGAGTCCAATT	987
Db	89617	CATTATCAAGGAAAGAGAGTCCAATT	89643
RESULT	5		
AC003682			
LOCUS			
DEFINITION		153875 bp DNA linear PRI 29-JUN-1998	
AC003682			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
cosfield,J., Duarte,S., Lucas,S., Bruce,R., Thomas,P., Quan,G., Kronmiller,B., Arellano,A., Montgomery,M., Ow,D., Nolan,M., Trong,S., Kobayashi,A., Olsen,A.S. and Carrano,A.V.			
Sequence analysis of a 2 Mb region containing a zinc finger (ZNF) gene cluster in 19q13.4			
Unpublished			
2 (bases 1 to 153875)			
Lamerdin,J.E.			
Direct Submission			
Submitted (16-DEC-1997)			
National Laboratory, 7000 East Ave., Livermore, CA 94551, USA			
3 (bases 1 to 153875)			
Lamerdin,J.E.			
Direct Submission			
Submitted (29-JUN-1998)			
Joint Genome Institute, Lawrence Livermore National Laboratory, 7000 East Ave., Livermore, CA 94551, USA			
On Jun 29, 1998 this sequence version replaced gi:2689440.			
Map and sequence oriented from q centromere to telomere.			
This accession derived from cosmid F18547 from bases 1 to 37,736, cosmid F11133 from bases 36,861 to 40,261, cosmid R27945 from bases 40,194 to 77,593, cosmid R28830 from bases 74,683 to 118,033, and cosmid R32804 from bases 114,809 to 153,875. This accession overlaps cosmid F25419 (AC003005) to the left from bases 1 to 5,157, and overlaps cosmid F20191 (AC004017) from bases 150,049 to 153,875. No sequence discrepancies were identified in any of the overlapping regions between cosmids.			
Additional chr 19 map and sequence information are available at: http://www-bio.lnl.gov/bbrp/genome/genome.html.			
Location/Qualifiers			
1. 153875			
/organism="Homo sapiens"			
/db_xref="taxon:9606"			
/chromosome="19"			
/map="19q13.4 between D19S773 to ZNF134"			
/clone="Splice of sequence from cosmids F18547-F11133-R27945-R28830-R32804"			
/cell_lines="UV5HL9-5B for F library clones, 5HL2-B for R"			
/note="Cosmid libraries LL19NC02 and LL19NC03 were constructed at LNL from flow-sorted chromosomes from human-hamster hybrids UV5HL9-5B and 5HL2-B, respectively, which carry chromosome 19 as their only human chromosome."			
complement(1..62)			
/rpt_family="Alu"			
101..248			
/rpt_family="L1"			
247..808			
/rpt_family="L1MB8"			
813..1114			
/rpt_family="TIGGER2"			
complement(1115..1325)			
/rpt_family="AluSc"			
1329..1752			
/rpt_family="LTR3"			
2245..4769			
/rpt_family="HERVK"			
6786..7215			
/rpt_family="LTR3"			
complement(7217..7296)			
/rpt_family="AluY"			
7518..7794			
/rpt_family="AluJo"			
complement(7795..7919)			
/rpt_family="AluJo/FLAM"			
7923..8719			
/rpt_family="TIGGER2"			
complement(8723..9023)			
/rpt_family="AluSg"			
9024..9489			
/rpt_family="TIGGER2"			
9498..9529			
/rpt_family="MER28"			
9530..9607			



Db 5171 ACCAGGACCCCTAACTCCCTGACAGAGAGTGAATGATCTTCAGAACAAATGGTATCAA 5230

Qy 480 TTATACCACTCCATGGAGGACTTCCTTTATGTGTCCACCAGGATACATGCTCAACTG 539

Db 5231 TTATATGCACCATGGAGGACTT-CITTTATGTGTCCACCAGGATGCATCACTCAACCG 5289

Qy 540 CAGTTGCTTCAGTTGATCCCAAGCATGGTTGAGTTACCAATAAAAAATATGTACCT 599

Db 5290 CAGTTGCTTCAGTTGATCCCAAGCATGGTTGAGTTACCAATAAAAAATATGTACCT 5349

Qy 600 ATTAGACCTTACCTTTATTAATATTTACTGTGTAGTTACTAATCACTCCTGGCCCATCA 659

Db 5350 ATTAGGCTTACCTTATTAATATTTACTGTGTAGTTACTAATCACTCCTGGCCCATCA 5409

Qy 660 CCCAAATGTACTGATTATACAGAAATGGGCTCCCTTTTGATTAATCTCACCCCTCCCTTG 719

Db 5410 CCCAAATGTATTGATTGTACAGAATGGGCTCCCTTTTGATTAATCTCACCCCTGCTCTG 5469

Qy 720 GGCCCACTGCTTGGCCCTTAGCTAGACAAATAGTCCATGTTAATGGGAGACATATTGA 779

Db 5470 GACTCAGTGTCTTGGCCCTTAGCTAACAACAGTCCATGTTAATGGGAGACATATTGA 5529

Qy 780 CTGGGTCCTGCTGCTCATTAAAGATGGAGAGATGAGAATCAGACACATGSCATAACT 839

Db 5530 CTGGGTCCTGCTGCTCATTAGATGGAGAGATGAGAATCAGACCTCATGGCATAATT 5589

Qy 840 TCACTGGCACTGGTGGGAACTTTAATCATCTTCACTTCAACACACTGGGATTCATC 899

Db 5590 TTGCTGCACTGGTGGGAACTTTAATCATCTTCACTTCAACACACTGGGATTCATC 5633

Qy 900 CCAATCTGCCATGCACTGCTGGCATGGAACGGGCTTTAGCCCACTTTGCTCCTCAATG 959

Db 5634 CCAATCTGCTACACAACCTGCTGGCAGAGATGGGCTTTAGCCCACTTTGCTCCTCAATG 5693

Qy 960 GCATTATCAAGGAAGAGAGTCCAAATTCAGGAGTCTATGTG 1001

Db 5694 GCATTATCAAGGAAGAGAGTCCAAATTCAGGAGTCAATATG 5735

RESULT 6

AC010642 45383 bp DNA linear PRI 26-AUG-2000

LOCUS

DEFINITION Homo sapiens chromosome 19 clone LUNLR-245B6, complete sequence.

ACCESSION AC010642

VERSION AC010642.5 GI:9929687

KEYWORDS HTG.

SOURCE Homo sapiens.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 45383)

DOE Joint Genome Institute and Stanford Human Genome Center.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 45383)

DOE Joint Genome Institute.

AUTHORS Direct Submission

TITLE Submitted (16-SEP-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

JOURNAL

REFERENCE 3 (bases 1 to 45383)

DOE Joint Genome Institute and Stanford Human Genome Center.

AUTHORS Direct Submission

TITLE Submitted (26-AUG-2000) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

JOURNAL

COMMENT On Aug 26, 2000 this sequence version replaced gi:7711487.

Draft Sequence Produced by DOE Joint Genome Institute

www.jgi.doe.gov

Finishing Completed at Stanford Human Genome Center

www-shgc.stanford.edu

Quality: Phrap Quality >=40 99.5% of Sequence;

Estimated Total Number of Errors is 0.3.

SVS Content:

SHGC-13293 G13764

WI-3037 G04140.

Location/Qualifiers

source 1. 45383

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="19"

/clone="LUNLR-245B6"

BASE COUNT 12088 a 11041 c 11063 g 11191 t

ORIGIN

Query Match 64.9% Score 649.2; DB 9; Length 45383;

Best Local Similarity 81.0%; Pred. No. 1.4e-190;

Matches 81; Conservative 0; Mismatches 168; Indels 23; Gaps 4;

Qy 1 GGAGATGATAAAGCGGTGTGAGTGCCTCAAGTTGTGTGCGACCATGGAATGGGAGACTG 60

Db 8105 GGAGATGATAAAGCGGTGTGAGTGCCTCAAGTTGTGTGCGACCATGGAATGGGAGACTG 8164

Qy 61 GAGGATATACATGATCCCAACTACAGCCCGAGCTCCTCCAGTATGAGCCATGAGCCAGTT 120

Db 8165 GAGGATATACATGATCCCAACTACAGCCCGAGCTCCTCCAGTATGAGCCATGAGCCAGCT 8224

Qy 121 GAATCTGAATGAGAGATGGAATGAAGACCGACGAGAGTGCACACTGAGTCAACCCCTCAT 180

Db 8225 GAATCTGAATGAGAGATGGAATGAAGACCGACGAGAGTGCACACTGAGTCAACCCCTCAT 8284

Qy 181 AACATGGGCTCAGATCAAGAAAACACACAGAAAGCTGAGAAACTGGTGTAGTGCACGGG 240

Db 8285 AACATGGGCTCAGATCAAGAAAACACACAGAAAGCTGAGAAACTGGTGTAGTGCACGGG 8344

Qy 241 TCAGGCAAAAACCCCTGACTCCATGTTTATGAGCCATGCTAGCTGTAATATCCTGTGCAGT 300

Db 8345 TCAGGCAAAAACCCCTGACTCCATGTTTATGAGCCATGCTAGCTGTAATATCCTGTGCAGT 8404

Qy 301 ATGATTTTCTGTGCGAGAGCAAAAACATATTTGGGCATATTTTCTTAACCCACCGGTAGT 360

Db 8405 ATGATTTTCTGTGCGAGAGCAAAAACATATTTGGGCATATTTTCTTAACCCACCGGTAGT 8464

Qy 361 GTGA-TCATACTCTGAAGCAGCACTCCTCTGAGATATATCATGATCAAGGAGCATCAGT 419

Db 8465 ACAGTGGTACTCTGGAGGACACTCCTCTGAGATATATTTATGATCAGGAGGCGTGGG 8524

Qy 420 ACCAGGACCTCTAATCCCTCGACAGAGCAATTAAGACTCTCATACAAATGATATCAA 479

Db 8525 ACCAGGACCTCTAATCCCTCGACAGAGCAATTAAGACTCTCATACAAATGATATCAA 8567

Qy 480 TTATACCACTCCATTTGAGGAGGACTTCCTTTATGTGTCCACCAGGATACATGCTCAACTG 539

Db 8568 TTATACCACTCCATTTGAGGAGGACTTCCTTTATGTGTCCACCAGGATACATGCTCAACTG 8627

Qy 540 CAGTTGCTTCAGTTGATCCCAAGCATGGTTGAGTTACCAATAAAAAATATGTACCT 599

Db 8628 CAGTTGCTTCAGTTGATCCCAAGCATGGTTGAGTTACCAATAAAAAATATGTATTT 8687

Qy 600 ATTAGACCTTACCTTTATTAATATTTACTGTGTAGTTACTAATCACTCCTGGCCCATCA 659

Db 8688 ATTAGACCTTACCTTTATTAATATTTACTGTGTAGTTACTAATCACTCCTGGCCCATCA 8747

Qy 560 CCCAAATGTACTGATTATACAGAAATGGGCTCCCTTTTGATTAATCTCACCCCTCCCTTG 719

Db 8748 CCCAAATGTACTGATTATACAGAAATGGGCTCCCTTTTGATTAATCTCACCCCTCCCTTG 8807

Qy 720 GGCCCACTGCTTGGCCCTTAGCTAGACAAATAGTCCATGTTAATGGGAGACATATTGA 779

Db 8808 GGCCCACTGCTTGGCCCTTAGCTAGACAAATAGTCCATGTTAATGGGAGACTTATTTGA 8864

Qy 780 CTGGGTCCTGCTGCTCATTAAAGATGGGAGAGATGAGAATCAGACACATGSCATAACT 839

Db 8865 CTGGGTCCTGCTGCTCATTAAAGATGGGAGAGATGAGAATCAGACCTCTTGTGGCAAACT 8924

Qy 840 TCACTGGCACTGGTGGGAACTTTAATCATCTTCACTTCAACACACTGGGATTCAT- 898

Db 8925 TCACTGGCACTGGTGGGAACTTTAATCATCTTCACTTCAACACACTGGGATTCAT- 8984

```
Qy 899 -CCCAATCTGCCATGCAACTTGTCTTGGCATGGAAGGGCTTTAGCCCACTTTGCTGCTCAA 957
      ||| || ||||| ||| || ||||| ||| || |||||
Db 8985 CCCAGTTGGCTGCACAACCTTGTGTATAGGACAGCGCTTTAAACCCGCCCTTACCTCAG 9044
      ||| || ||||| ||| || ||||| ||| || |||||
Qy 958 TGGCATTTATCAAGGAAAGAGAGGTCCTCAATTCAGAGTCTATGTG 1001
      ||| || ||||| ||| || ||||| ||| || |||||
Db 9045 TGACGTCATCTAGGAACAGAGGCCCAATTTCAGGAATCGATATG 9088
      ||| || ||||| ||| || ||||| ||| || |||||

RESULT 7
AC020915 160852 bp DNA linear PRI 31-MAY-2001
DEFINITION Homo sapiens chromosome 19 clone CTD-3138B18, complete sequence.
ACCESSION AC020915
VERSION AC020915.6 GI:14269683
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 160852)
TITLE DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Direct Submission
AUTHORS Unpublished
REFERENCE 2 (bases 1 to 160852)
TITLE DOE Joint Genome Institute.
JOURNAL Direct Submission
AUTHORS Submitted (12-JAN-2000) Production Sequencing Facility, DOE Joint
JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
AUTHORS 3 (bases 1 to 160852)
TITLE DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Direct Submission
AUTHORS Submitted (31-MAY-2001) DOE Joint Genome Institute, 2800 Mitchell
JOURNAL Drive, Walnut Creek, CA 94598, USA
COMMENT On May 31, 2001 this sequence version replaced gi:13752690.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.9% of Sequence;
Estimated Total Number of Errors is 0.1.
STS Content:
SHGC-9787 G11320.

FEATURES
source Location/Qualifiers
1. 160852
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="CTD-3138B18"
BASE COUNT 42283 a 37137 c 39005 g 42427 t
ORIGIN

Query Match 64.9%; Score 649.2; DB 9; Length 160852;
Best Local Similarity 81.0%; Pred. No. 1.7e-190;
Matches 813; Conservative 0; Mismatches 168; Indels 23; Gaps 4;

Qy 1 GGAGATGGATAAACCGTGTGAGTGGCCCTCAAGTCTGTGCGACCATGGAATGGGAGACTG 60
      ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 159499 GGAGATGGACAACCGTGTGGGTGGCCCTCCAGGTGTGTGTACCATGGAAACGGGAGACCG 159558
      ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Qy 61 GAGGATACATGGATCCCAACTACAGGCCCGCAGCTCCTCCAGTATGAGCCATGAGCCAGTT 120
      ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 159559 GAGGATCCATGGTATTCAACCGTGGCGCTGTATTCTCCAGTACGACCATGAGCCAGCG 159618
      ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Qy 121 GAATCTGAATGGATGGATGAATGAAGACCGACGAGTGCACATGACATCAACCTCAT 180
      ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 159619 GAATCTGAATGAAGACGACAGACAGGCGCCAGCGAGTCAACAATGACATCAACCCCAT 159678
      ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Qy 181 AACATGGGTGATCAAGAAACACACACAGCAAGCTGAGAACTGGTGTAGTGCAGGG 240
      ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 159679 AACATGGGACAGATCAAGAAACACACACAGCAAGCTGAGAACTGAGTGGAGCGCCAGGG 159738
      ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
```





Db 44001 ACCAGAGCCCTTAACCTCCGCTGCACAGAG-----CAATAGCATCAA 44043  
Qy 480 TTATACCACTCCATGAGGAGCTCCCTTTATGTGTACCCAGGATACATTCCTCAACTG 539  
Db 44044 TTACACCGGCCCATTTGGAGGACTCCCTTTATGTATACCATGGATAGTCACTCACTG 44103  
Qy 540 CAGTTGCCCTTGACGTTTGTCCCAAGCATGGTTAGTTACCATAAAAAAATATGTACTT 599  
Db 44104 CAGCTGTCTTGACGTTCAATCTCAGGCATGGTTAAGTCACCATGGAAAAAATATGTATT 44163  
Qy 600 ATTACACCTTAGCTTTATTAATATATACCTTGCTAGTACTACTATCACTCCTGGCCCCATCA 659  
Db 44164 AFTAGGCTCAGCTTTATTAATGTTACGGGCTGTTTACCACTACTCCTAGCCCCATCA 44223  
Qy 660 CCCAAATGTACTGATTATACAGAATGGCTCCCTTTGATATCTCACCCCTCCCTG 719  
Db 44224 CTCAAATGTATGATATATGATGATGGCTCCCTTTGATATCTCACCCCTCCCTG 44283  
Qy 720 GCCCACCCTGCTTGGCCCTTAGCTAGACATATAGTCCATGTTAATGGGAGACATTTATGA 779  
Db 44284 GACCAGTGTCTTGGCCCTTGGCCAGACACAGTCTATGTTAATGGGAAC---ATTGA 44340  
Qy 780 CTGGGGTCCCTGCTGCTATTAGATGGGAGAGATGAGATCAGACCATGGCATAAAT 839  
Db 44341 CTGGGGTCCCTGCTGCTATTAGATGGGAGAGATGAGATCAGACCATGGCATAAAT 44400  
Qy 840 TCATCGGCACTGGTGGCGAACTTTAATCACTCTTCACTTCAACACACTGGGATTCAT- 898  
Db 44401 TCATCGGCACTGGTGGCGAACTTTCAACCTCTTCACTTCAACACACTGGGATTCATC 44460  
Qy 899 -CCCAATCTGCCATCAACTTCTTGGCATGGAACGGCTTTAGCCCACTTTGGCTCAA 957  
Db 44461 CCCAGTGGCTGCACAACTTCTGTTATAGGACAGGCTTTAAACCCGCCCTTACCTCAG 44520  
Qy 958 TGGCATTATCAGGAAGAGAGAGGTCATTCAGGAGTCTATGTG 1001  
Db 44521 TCAGCTCATCTAGGAACAGAGGCGCCCAATTCAGGAATCGATATG 44564

RESULT 9  
AC015589  
LOCUS  
DEFINITION Homo sapiens clone RP11-44L20, WORKING DRAFT SEQUENCE, 32 unordered pieces.  
AC015589  
VERSION AC015589.4 GI:10047664  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 158802)  
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
TITLE Homo sapiens chromosome, clone RP11-44L20  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 158802)  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barna,N., Beckerly,R., Boguslavskiy,L., Boukhalter,B., Brown,A., Castle,A., Colangelo,M., Collins,S., Collumore,A., Cooke,P., Dearlilano,K., Dewar,K., Domino,M., Doneilan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,G., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lehoczy,J., Lien,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye.W.J., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

## COMMENT

On Sep 10, 2000 this sequence version replaced gi:7107718.  
All repeats were identified using RepeatMasker:  
Snit, A.P.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L1096  
Center clone name: 44\_L20  
----- Summary Statistics  
Sequencing vector: M13; M7815; 100% of reads  
Chemistry: Dye-primer-amersham; 6% of reads  
Chemistry: Dye-terminator Big Dye; 94% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 139045 bases at least Q40  
Consensus quality: 146903 bases at least Q30  
Consensus quality: 150628 bases at least Q20  
Insert size: 170000; agarose-fp  
Quality coverage: 3.2 in Q20 bases; agarose-fp  
Quality coverage: 3.5 in Q20 base.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 32 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 5488: contig of 5488 bp in length  
\* 5489 5588: gap of 100 bp  
\* 5589 7161: contig of 1573 bp in length  
\* 7162 7261: gap of 100 bp  
\* 7262 8657: contig of 1396 bp in length  
\* 8658 8757: gap of 100 bp  
\* 8758 10159: contig of 1402 bp in length  
\* 10160 10259: gap of 100 bp  
\* 10260 12255: contig of 2000 bp in length  
\* 12260 12359: gap of 100 bp  
\* 12360 14250: contig of 1891 bp in length  
\* 14251 14350: gap of 100 bp  
\* 14351 16194: contig of 1844 bp in length  
\* 16195 16294: gap of 100 bp  
\* 16295 17825: contig of 1531 bp in length  
\* 17826 17925: gap of 100 bp  
\* 17926 19224: contig of 1299 bp in length  
\* 19225 19324: gap of 100 bp  
\* 19325 21412: contig of 2088 bp in length  
\* 21413 21512: gap of 100 bp  
\* 21513 23847: contig of 2335 bp in length  
\* 23848 23947: gap of 100 bp  
\* 23948 25938: contig of 1991 bp in length  
\* 25939 26038: gap of 100 bp  
\* 26039 28260: contig of 2222 bp in length  
\* 28261 28360: gap of 100 bp  
\* 28361 31500: contig of 3140 bp in length  
\* 31501 31600: gap of 100 bp  
\* 31601 34208: contig of 2608 bp in length  
\* 34209 34308: gap of 100 bp  
\* 34309 37683: contig of 3375 bp in length  
\* 37684 37783: gap of 100 bp  
\* 37784 42304: contig of 4521 bp in length  
\* 42305 42404: gap of 100 bp  
\* 42405 46675: contig of 4271 bp in length  
\* 46676 46775: gap of 100 bp  
\* 46776 60170: contig of 1395 bp in length  
\* 60171 60270: gap of 100 bp  
\* 60271 64806: contig of 4536 bp in length  
\* 64807 64906: gap of 100 bp  
\* 64907 69745: contig of 4839 bp in length



Db 110722 CCAAAATTGATTATATACAGATGGCTCCCTTTGATAGTTCTACCACCCCTCC 110781

Qy 717 TTGGGCCCACTGCTGTGGCCCTTAGCTAGACAATFAGTCCATGTGTAATGGGAGACATTAT 776

Db 110782 TTGGACCAATGTCTGTGGCTTTAGCTAGACAACAGTCTATGCTAATGGGAGACATTGG 110841

Qy 777 TGACTGGGGTCCCTGTGTCATTAGATGGGAGAGATGAGATCAGACCATGGCATAA 836

Db 110842 TGACTGGGGACCCCGTGTATTAATGATGGAAGAGATGAGATCAGACCTCATGGSCAAA 110901

Qy 837 ACTTCACATGGCAGCTGGTGGCGAACTTTAACATCTCTTCACTTCAACACACTGGGATTCA 896

Db 110902 ACTTCGCTGGCAGCTGGTGGCAGACCTTCAACATCTCTTCACTACATCAGACTGGGATTCA 110961

Qy 897 ATCCCAATCTGCCATCAACTGCTTGGCATGGAACGGGCTTTAGCCACACCTTTGCCATCA 956

Db 110962 ATCCCAAGTCGGCTGTGGCTTGTGGCAGACAGCAGGCTTAGCTCGCCTTGGCCTCA 111021

Qy 957 ATGGCATTTATCAAGGAAGAGAGGTCCTCAATTCAGGAGTCATGTG 1001

Db 111022 GTGGCAGCATCAAGGAAGAGGGGCCCAATTCAGGAGTCGATATG 111066

RESULT 10

AC010328 196686 bp DNA linear PRI 13-JUL-2002

LOCUS Homo sapiens chromosome 19 clone CITB-EL\_2620122, complete

DEFINITION

ACCESSION AC010328

VERSION AC010328.4 GI:7381733

KEYWORDS htg..

SOURCE

ORGANISM Homo sapiens

REFERENCE

AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

TITLE 1 (bases 1 to 196686)

JOURNAL DOE Joint Genome Institute and Stanford Human Genome Center.

REFERENCE

AUTHORS Direct Submission

TITLE 2 (bases 1 to 196686)

JOURNAL DOE Joint Genome Institute.

REFERENCE

AUTHORS Direct Submission

TITLE 3 (bases 1 to 196686)

JOURNAL Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE

AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.

TITLE Direct Submission

JOURNAL Submitted (01-APR-2000) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE

AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.

TITLE Direct Submission

JOURNAL Submitted (18-APR-2000) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE

AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.

TITLE Direct Submission

JOURNAL Submitted (13-JUL-2002) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT

On Apr 1, 2000 this sequence version replaced gi:6997076.

Draft Sequence Produced by DOE Joint Genome Institute

www.jgi.doe.gov

Finishing Completed at Stanford Human Genome Center

Quality: Phrap Quality >=40 99.5% of Sequence;

Estimated Total Number of Errors is 0.9.

STS Content:

SHGC-32227 G31446.

Location/Qualifiers

1..196686

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="19"

FEATURES

source

BASE COUNT 50946 a 43545 c 45142 g 57053 t

ORIGIN

Query Match 62.5%; Score 625.8; DB 9; Length 196686;

Best Local Similarity 79.0%; Pred. No. 3.4e-183;

Matches 794; Conservative 0; Mismatches 182; Indels 29; Gaps 3;

Qy 1 GGAGATGGATAAACCGTGTGAGTGGCCCTCAAGTTGTGTGCGACCATGGAATGGGAGACTG 60

Db 159354 GGAGATGGACAACCATCTGTGGTGGCCCTCCAGGTGTGTGCGACCATGAATGGGAGACTA 159413

Qy 61 GAGGATACATGGATCCCACTACAGGCCAGCTCCTCCAGCATATGAGCCATGAGCCAGTT 120

Db 159414 GCAGGACCCACGTTGGCCCAACCGTGGACCTGGTCCCTCCAGTAGGAGTCGTGAGCCAGCT 159473

Qy 121 GAATCTGAATCTGAAGATGGAATGAAGACCCAGCAGAGTGCACACTGCAGCTCAACCCCTCAT 180

Db 159474 GAATTTGAATGCAAGATGGAGCAAGGCCCGCACTGGAGTGCACACTGCACATCCACCCCAT 159533

Qy 181 AACATGGGGTCCAGATCAAGAAAAACACACAGAAAGCTTGAGAACTGGTGTAGTGCCAGGG 240

Db 159534 AACATGGGGCAGATCAAGAAAAACAAAGCAGGAAGCTGAGAAAAATGCTGGA----- 159584

Qy 241 TCAGGCAAAAACCCCTGACTCCATGTTTATGGCCATGCTAGCTGTAATATCCTGTGCAGT 300

Db 159585 -----GCATCATGTGCTTGGCCATGTTAGCAGTAGTGTCTGTGAGGT 159628

Qy 301 ATGATTTTCTGTCAGAAAGCAAAAACATATTGGGCATATTTTCTTAACCCACCGGTAGT 360

Db 159629 ATGTTTCCCTCTGTAGAGGCCAAACACATCTGGGCATATGTTCCTAATCACCCTGTAGT 159688

Qy 361 GTGATC-ATACTCTGAAGCAGCAGCTCCTCCTGAGATATATCATGATCAAGGAGCATCAT 419

Db 159689 GTGACCTGTACTCTGGAGTGACATTCTCTCTGAGGTATATCATGATCAGGAGCGTGGGC 159748

Qy 420 ACCAGGACCTTAACCTCCCTCGACAGACAACAATTAGACTCTCATACAATGTTATCAA 479

Db 159749 TCTAGGGCCCTAACTCCCTCGATATAGAGCAATTAACACTCTCACAACAATAGATCAG 159808

Qy 480 TTATACCACTCCATTTGGAGGAGCTTCTCTTTATGTCTACCCAGGATACATATGCTCAACTG 539

Db 159809 TTATACGCCCTCATTTGGAGGAGCTCCCTTTATGATCACCACCAAGATATGTCACACA 159868

Qy 540 CAGTTGCCCTTGCAAGTTGATCCCAAGCATGGTGTAGTTTACCATAAAAAATATATGTACT 599

Db 159869 CAGTTGTCTTGCAAGTCCAAACCCAGGCATGGTTGAGTCACCATGGAAAAGTTATGTATTT 159928

Qy 600 ATTAGACCTTAGCTTTATTAATATTAATTTACTTGTGTAGTTACTACTCTGCCCCCATCA 659

Db 159929 ATTAGGCCCTTAGCTCTATTAATAAATAGTTGTGCTCATCAACCACTCCCGCCCCCATCA 159988

Qy 660 CCCAAATGTACTGATATACAGAAATGGCTCCCTTTGATTAATTTCT---CACCCCCCTCC 716

Db 159989 CCAAAATTTGATTGATATACAGAAATGGCTCCCTTTGATAGTTCTCACCACCCCTCC 160048

Qy 717 TTGGGCCCATGCTCTTGCCCTTTAGCTAGACAATFAGTCCATGTGTAATGGGAGACATTAT 776

Db 160049 TTGGACCAATGTCTTGTGCTCTTTAGCTAGACAACAGTCTATGCTAATGGGAGACATTGG 160108

Qy 777 TGACTGGGGTCCCTGTGTCATTAGATGGGAGAGATGAGATCAGACCATGGCATAA 836

Db 160109 TGACTGGGGACCCCGTGGTTAATTAGATGGAAGAGATGAGATCAGACCTCATGGCAAA 160168

Qy 837 ACTTCACATGGCAGCTGGTGGCGAACTTTAACATCTCTTCACTTCAACACACTGGGATTCA 896

Db 160169 ACTTCGCTGGCAGCTGGTGGCAGACCTTCAACATCTCTTCACTACATCAGACTGGGATTCA 160228

Qy 897 ATCCCAATCTGCCATGCAACTGCTGTGCGATGGAACGGGCTTTAGCCACACCTTTGCCCTCA 956

Db 160229 ATCCCAAGTCGGCTGTGCGGCTTGTCTTGGCAGACAGCAGGCTTTAGCTCGCCTTTGCCCTCA 160288

Qy 957 ATGGCATTTATCAAGGAAGAGAGGTCCTCAATTCAGGAGTCTATGTG 1001



```
repeat_region /rpt_family="Alu"
7359..7481
repeat_region /rpt_family="MaLR"
7482..7577
repeat_region /rpt_family="HY3"
7583..7627
repeat_region /rpt_family="(TAAA)n"
7628..7663
repeat_region /rpt_family="MaLR"
7777..8072
repeat_region /rpt_family="Alu"
8144..8252
repeat_region /rpt_family="L1"
8250..8265
misc_feature /note="match to EST BM015942 (NID:g16530296),"
8253..8549
repeat_region /rpt_family="Alu"
8550..9057
repeat_region /rpt_family="L1"
9058..9181
repeat_region /rpt_family="Alu"
9112..9405
misc_feature /note="match to EST BG719345 (NID:g13998532),"
9182..9279
repeat_region /rpt_family="L1"
9280..9604
repeat_region /rpt_family="Alu"
9605..9693
repeat_region /rpt_family="L1"
9694..10001
repeat_region /rpt_family="Alu"
10002..10356
repeat_region /rpt_family="L1"
10464..10832
misc_feature /note="similar to Sus scrofa EST AW479624 (NID:g7049730),"
11011..11093
repeat_region /rpt_family="MIR"
11312..11638
repeat_region /rpt_family="Alu"
11320..11335
misc_feature /note="match to EST AW022240 (NID:g5875770) df34g09.yl"
11322..11335
misc_feature /note="match to EST AI207473 (NID:g6361481)"
11845..11851
misc_feature /note="match to EST BE294168 (NID:g9177699),"
12037..12101
repeat_region /rpt_family="Alu"
12558..12623
misc_feature /note="match to EST AA010360 (NID:g1471396) z108f11.sl"
12382..13251
misc_feature /note="CpG island (%GC=65.6, o/e=0.80, #CpGs=77)"
12488..12751
misc_feature /note="match to EST BG719345 (NID:g13998532)"
12759..13292
misc_feature /note="match to EST BG720169 (NID:g13999356)"
13168
misc_feature /note="match to EST AA010359 (NID:g1471395) z108f11.rl"
13213..13292
misc_feature /note="match to EST AA010359 (NID:g1471395) z108f11.rl"
13277..13330
```

Query Match 61.4%; Score 615; DB 9; Length 124102;  
Best Local Similarity 86.7%; Pred. No. 7.4e-180;  
Matches 703; Conservative 0; Mismatches 100; Indels 8; Gaps 2;

```
QY 191 CAGATCAGAAACACACACAGAGCTGAGAACTGGTGTAGTCCAGGGTCAGGCAGAAA 250
Db 107348 CAGATCAGAAACACACAGAGATGAGAACTGCTGGAGCATCAGGGTCAGGCAGAAA 107289
QY 251 ACCCTGACTCCATCTTTATGCCATGCTAGCTGTGAATATCTCTGTGACAGTATGATTTTC 310
Db 107288 ACCCTGACTCCATGTTTATGGCCATGCTAGCTATGATATCTCTGTGCTATGTTTTCCC 107229
```

```
QY 311 TGTGAGAGCAAAACATATTTGGGCATATTTTCTTAACCCACCGGTAGTGTGATCATPAC 370
Db 107228 TGTGAGAGGCAAAACATATTTGGGCTTATGTTCTTAACCCCGGTAGTATGA ----- 107175
QY 371 TCTGAGCAGCACTCCCTCTGAGATATATCATGATCAAGAGGACATCAGTACGAGCACCTC 430
Db 107174 -CCGGGGTGACACTCCCTCTGAGATATATCATGATCAAGAGGACATGCGACCGGACCTC 107116
QY 431 TAACTCCCTCTGACACAGAGCAATTAGACTCTCTAACAATGGTATCAATTTATACCACTC 490
Db 107115 TAACTCCCTCTGACACAGAGCAATTAGACTCTCTCAGAACAAATGGTATCAATTTATACCACTC 107056
QY 491 CAITGGAGGGACTTCCTTTATGTGTCAACCCAGGATACATTTGCTCAACATGCAAGTTCCTTGC 550
Db 107055 CATTGGAGGGCTTTCTTTATGTCATCAACCCAGGATACATTTGCTCAACATGCAAGTTCCTTGC 106996
QY 551 CAGTTTGCATCCCAAGCAGTGTGAGTTACCATATAAAAAATTTATGTACCTATTAGACCTTA 610
Db 106995 CAATTCAATCCCGAGCATGTTGAGTTACCATTGGAATAATTTACGTACCTATTGGGCCCTTA 106936
QY 611 GCTTTATTAAATTACTTGTGTAGTTACTAATCACTCTCTGCGCCCATCACCCAAATTTGTA 670
Db 106935 GTTCTACTAATATTACTGGTGTAGTTACTAATCACTCTCTGCGCCCATCACCCAAATTTGTA 106876
QY 671 CTGATTATACAGAAATGGGCTCCCTTTGATAATCTCAACCCCTCTCTTGGGCCCACTGTC 730
Db 106875 TTAATTATATAGAATGGGCTCCCTTTGATAATCTCAACCCCTCTCTTGGGCCCACTGTC 106816
QY 731 TTGCCCCCTTAGCTAGACAAATAGTCCATGTTTAATGGGAGACATTTATGACATGGGGTCCCT 790
Db 106815 TTAGTCCCTTAGCTAGACAAACAGTCCATGCTTAATGGGAGACATTTATGACATGGGGTCCCT 106756
QY 791 GTGTGCTATTAAAGATGGGAGATGAGATCAGACATGAGACATGACATAAACTTCACCTGGCACT 850
Db 106755 ATGTCATTTGTTGGGAGACATGAGCATCAGACCTCATGCGATTAATTTGCTGGAAT 106696
QY 851 GGTGGCAAACTTTAACAATCTCTTCACTTCAACACATCGGGATTCAATCCCAATCTGCCA 910
Db 106695 GGTGCAAACTTTAACAATCTC-CCACTACATCAGATTGCAATTCATCTCAATCTGCTG 106637
QY 911 TGCAACTGCTTGTCATGGAACGGCTTTAGCCACCTTTGCCCAATGGCATTTATCAAG 970
Db 106636 CCCAAATTTGTCAGACAGCATGCTTTAGCTACCTTTGCTCAATGGCATTTATCAAG 106577
QY 971 GAAAGAGAGGTCCTCAATTCAGGAGTCTATGTG 1001
Db 106576 GAAAGAGAGGTCCTCAATTCAGGAGTCAATATG 106546
```

```
RESULT 12
AC069481 158302 bp DNA linear HTG 24-AUG-2002
LOCUS Homo sapiens chromosome 5 clone RP11-153J19 map 5, WORKING DRAFT
DEFINITION SEQUENCE, 28 unordered pieces.
ACCESSION AC069481
VERSION AC069481.2 GI:10567949
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 158302)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 5, clone RP11-153J19
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 158302)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Cullymore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
```

Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
 Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,  
 Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J.,  
 Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,  
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,  
 Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J.,  
 Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,  
 O'Neil, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,  
 Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
 Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J.,  
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,  
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.

# TITLE JOURNAL

Submitted (01-JUN-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 158302)

# REFERENCE AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
 Anderson, S., Barna, N., Bastien, V., Bida, F., Boguslavsky, L.,  
 Boukhgalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A.,  
 Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,  
 Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Ferreira, P.,  
 FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M.,  
 Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L.,  
 Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., LaRocque, K.,  
 Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G.,  
 Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McKernan, K.,  
 McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Mienga, V.,  
 Morrow, J., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T.,  
 O'Donnell, P., O'Neil, D., Oliver, T. M., Oliver, J., Peterson, K.,  
 Pierre, N., Pisan, C., Pollara, V., Raymond, C., Rieback, M., Riley, R.,  
 Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P.,  
 Sougnuez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
 Strausz, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
 Tirrell, A., Travers, M., Triggillo, J., Vassiliev, H., Viel, R., Vo, A.,  
 Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J.,  
 Zimmer, A. and Zody, M.

# TITLE JOURNAL

Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Oct 4, 2000 this sequence version replaced gi:8137154.

# COMMENT

All repeats were identified using RepeatMasker:  
 Smit, A. F. A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L10412

Center clone name: 153\_J\_19

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 145779 bases at least Q40

Consensus quality: 151559 bases at least Q30

Consensus quality: 153749 bases at least Q20

Insert size: 155602; sum-of-contigs

Quality coverage: 4.8 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 28 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 2578: contig of 2578 bp in length

\* 2579 2678: gap of 100 bp

\*

\* 2679 3836: contig of 1158 bp in length  
 \* 3837 3936: gap of 100 bp  
 \* 3937 5015: contig of 1079 bp in length  
 \* 5016 5115: gap of 100 bp  
 \* 5116 23885: contig of 18770 bp in length  
 \* 23886 23985: gap of 100 bp  
 \* 23986 25726: contig of 1741 bp in length  
 \* 25727 25826: gap of 100 bp  
 \* 25827 27337: contig of 1511 bp in length  
 \* 27338 27437: gap of 100 bp  
 \* 27438 29802: contig of 2365 bp in length  
 \* 29803 29902: gap of 100 bp  
 \* 29903 32473: contig of 2571 bp in length  
 \* 32474 32573: gap of 100 bp  
 \* 32574 35307: contig of 2734 bp in length  
 \* 35308 35407: gap of 100 bp  
 \* 35408 38588: contig of 3181 bp in length  
 \* 38589 38688: gap of 100 bp  
 \* 38689 41368: contig of 2680 bp in length  
 \* 41369 41468: gap of 100 bp  
 \* 41469 44056: contig of 2588 bp in length  
 \* 44057 44156: gap of 100 bp  
 \* 44157 48600: contig of 4444 bp in length  
 \* 48601 48700: gap of 100 bp  
 \* 48701 52600: contig of 3900 bp in length  
 \* 52601 52700: gap of 100 bp  
 \* 52701 58166: contig of 5466 bp in length  
 \* 58167 58266: gap of 100 bp  
 \* 58267 62266: contig of 4000 bp in length  
 \* 62267 62366: gap of 100 bp  
 \* 62367 65972: contig of 3606 bp in length  
 \* 65973 66072: gap of 100 bp  
 \* 66073 72761: contig of 6689 bp in length  
 \* 72762 72861: gap of 100 bp  
 \* 72862 79897: contig of 7036 bp in length  
 \* 79898 79997: gap of 100 bp  
 \* 79998 86476: contig of 6479 bp in length  
 \* 86477 86576: gap of 100 bp  
 \* 86577 94565: contig of 7989 bp in length  
 \* 94566 94665: gap of 100 bp  
 \* 94666 102190: contig of 7525 bp in length  
 \* 102191 102290: gap of 100 bp  
 \* 102291 111797: contig of 9507 bp in length  
 \* 111798 111897: gap of 100 bp  
 \* 111898 120738: contig of 8841 bp in length  
 \* 120739 120838: gap of 100 bp  
 \* 120839 130494: contig of 9656 bp in length  
 \* 130495 130594: gap of 100 bp  
 \* 130595 140773: contig of 10179 bp in length  
 \* 140774 140873: gap of 100 bp  
 \* 140874 157857: contig of 16984 bp in length  
 \* 157858 157957: gap of 100 bp  
 \* 157958 158302: contig of 345 bp in length.

# FEATURES

source

1. 158302  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="5"  
 /map="5"  
 /clone="RP11-153J19"  
 /clone\_lib="RPC1-11 Human Male BAC"

misc\_feature

1. 2578  
 /note="assembly\_fragment  
 clone\_end:SP6  
 vector\_side:left  
 2679. 3836  
 /note="assembly\_fragment"  
 3937. 5015  
 /note="assembly\_fragment"  
 5116. 23885  
 /note="assembly\_fragment"  
 23986. 25726  
 /note="assembly\_fragment"

misc\_feature

3937. 5015  
 /note="assembly\_fragment"

misc\_feature

5116. 23885  
 /note="assembly\_fragment"

misc\_feature

23986. 25726  
 /note="assembly\_fragment"





Query Match		60.1%;	Score 602;	DB 9;	Length 4473;
Best Local Similarity		80.5%;	Pred. No. 5.2e-176;		
Matches		811;	Conservative	0;	Mismatches 170; Indels 27; Gaps 8;
Qy	1	GGAGATGATTAACCGTGTAGTGGCCCTCAAGTGTGTGCGACCATGGAATGGGAGACTG	60		
Db	2150	GGAGTGACAAACCGTGTGGGTGCCCTCCAGGTGTGTGACCATGGAACGGAGACCG	2091		
Qy	61	GAGGATACATGGATCCCACTACAGCCGCCAGCTCCTCCAGTATGAGCCATGAGCCAGTT	120		
Db	2090	GAGGATCCATGGATTAACCGTGGCCCTGTACCTCCAGTACGAGCATGAGCCAGCG	2031		
Qy	121	GAATCTGAATGTGAAGTGAATGAAGACCGACGAGAGTACACTGACCTCAACCTCAT	180		
Db	2030	GAATCTGAATCAAAAGACAGAAAGGCGCCAGCGAGTCAATGACATCCAACCCCAT	1971		
Qy	181	AACATGGGGTCA-GATCAAGAAAACCAACACGAGAGCTGAGAACTGGTGTAGTGCCAGG	239		
Db	1970	AACATGGGGACAAGATCAAGAAAACCGACACAAGAGCTGAGAACTACTGGAGCGCCAGG	1911		
Qy	240	GTCAGGAAAAACCCCTGACTCCATGTTTATGGCCATGCTAGCTGTATATATCTGTGCG	299		
Db	1910	GTCAGCCTAAACCCCTGACTCCATGTTTATGGCCATGCTAGCTGTATATATCTGTGCGG	1851		
Qy	300	TATGATTTTCTGTGCGAAGCAAAAACATATTGGGCATATTTTCTTAACCCACCGGTAG	359		
Db	1850	TATGTTTCCCTGTGTAGAGCAAAAACATATTGGGCATATTTCTTAACCCACCGGTAG	1791		
Qy	360	TGTGA-TCATACTCTGAAGAGCACTCTCTCTGAGATATATCATGATCAAGAGCATCAG	418		
Db	1790	TACGACTGTCTACTGTGGAGCACACTCTCTCTGAGATATATATGATCAGGCGGTGGG	1731		
Qy	419	TACCAGGACCTTAACCTCCCTGCACAGAGCAATAGACTCTCATACAAATGGTATCA	478		
Db	1730	CACCAGGACCCCTAACTCCGCTGCACAGAG------CAATAGCATCA	1688		
Qy	479	ATTATACACATCCATTTGAGGAGACTTCCTTTATGTGTGCACCCAGGATACATTGCTCAACT	538		
Db	1687	ATTACACGGCCCATTTGGAAGCACTCCCTTTATGTATCATCACCATGATACGTCACCTCACT	1628		
Qy	539	CGAGTGGCTTGCAGTTGATCCCAAGCATGGTGTAGTTCACATAAATAATATGTATACC	598		
Db	1627	GCAGCTGTCTGTGACGTCAAACTCTCAGGCATGGTTAAGTTCACATGGAAAAATATGTATT	1568		
Qy	599	TATTAGACCTTAGCTTATTAATATTAATCTGTGTAGTTACTAATCACTCTGCGCCCATC	658		
Db	1567	TATTAGGCTCAGCTTATTAATGTACTGCGGTGTTCCACCAATCACTCTTAGCCCCCATC	1508		
Qy	659	-ACCCAATTTACTGATT-ATACAGAAATGGGCTCCCTTTTGATTAATCTCA-CCGCCCTC	715		
Db	1507	AACTCAAATTTATGGATTAATATGGAATGGCTCCCTTTTCATATTCTCACCCCTCC	1448		
Qy	716	CTTGGGCCCATGCTCTGTGGCCCTTTAGTAGACAATAGTGCATGTTAATGGGAGACATTA	775		
Db	1447	CTTGGACCATGCTCTGTGGCCCTTTGGCCAGACAACTGCTATGTTAATGGGAAAC---A	1391		
Qy	776	TTGACTGGGGTCCCTGTGGTCAATTAAGATGGAGAGATGAATCAGACCATGGCAT	835		
Db	1390	TTGACTGGGGTCCCGTGGCCATTTAGAGGGGAGATGAATAATCAGACCTCTTGGGCACA	1331		
Qy	836	AACCTCACTGGCACTGGTGGGAACTTTAAACATCTCTTCACTTCAACACACTGGGATTC	895		
Db	1330	AACCTCACTGGCACTGGTGGGACCTTCAACACTCTTCTCATCATCATCACTGGGATTC	1271		
Qy	896	AAT--CCCAATCTGCCATGCAACTGTGTGCGATGGAACGGCTTTAGCCCACTTTGCC	953		
Db	1270	AGTCCCGCAGTTGGCTGCACAACCTTGCTTGTGTATGAGACAGGCTTTAACCCGCCCTTACC	1211		
Qy	954	TCATGGCATTTATCAAGAAAGAGAGGTCCCAATTCAGGAGTCTATGTG	1001		
Db	1210	TCAGTGACGTATCTAGGAACAGAGGCCCAATTCAGGAATTCGATATG	1163		

RESULT 14	AL451075/c	AL451075	Human DNA sequence from clone RP11-215123 on chromosome 1, complete sequence.
LOCUS	AL451075	153937 bp	DNA linear PRI 26-MAY-2001
DEFINITION	AL451075	GI:14148854	
ACCESSION	AL451075		
VERSION	AL451075.15		
KEYWORDS	HTG.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	1	(bases 1 to 153937)	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	Coville,G.		
TITLE	Direct Submission		
JOURNAL	Submitted (26-MAY-2001)	Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk	
COMMENT	On May 18, 2001 this sequence version replaced gi:13751995. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at <a href="http://www.sanger.ac.uk/Projects/C_elegans/wormpep">http://www.sanger.ac.uk/Projects/C_elegans/wormpep</a> This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <a href="http://www.sanger.ac.uk/HGP/Chrl">http://www.sanger.ac.uk/HGP/Chrl</a> RP11-215123 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see <a href="http://www.chori.org/bacpac/home.htm">http://www.chori.org/bacpac/home.htm</a> VECTOR: pBAC3.6 This sequence is the entire insert of clone RP11-215123 The true left end of clone RP11-545A16 is at 133805 in this sequence. The true right end of clone RP4-632K5 is at 55227 in this sequence.		
FEATURES	Location/Qualifiers		
Source	1..153937	/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/chromosome="1"	
		/clone="RP11-215123"	
		/clone.lib="RPCI-11.1"	
	535..832		
repeat_region	/note="AluX repeat: matches 1..298 of consensus"		
repeat_region	1188..1484	/note="AluJo repeat: matches 1..289 of consensus"	
repeat_region	1834..2550	/note="L1MB8 repeat: matches 5346..6064 of consensus"	
repeat_region	2551..2851	/note="AluSx repeat: matches 2..300 of consensus"	
repeat_region	2852..2891	/note="L1MB8 repeat: matches 5307..5346 of consensus"	
repeat_region	2892..3198	/note="AluSx repeat: matches 1..310 of consensus"	
repeat_region	3199..3957	/note="L1MB8 repeat: matches 4474..5307 of consensus"	
repeat_region	4007..4177	/note="AluSg/x repeat: matches 140..312 of consensus"	
repeat_region	4185..4488		

```
repeat_region /note="AluSx repeat: matches 1. .304 of consensus"
repeat_region 4627. .4746
repeat_region /note="FLAM_C repeat: matches 7. .126 of consensus"
repeat_region 5157. .5466
repeat_region /note="AluJo repeat: matches 1. .309 of consensus"
repeat_region 5659. .5802
repeat_region /note="AluSg/x repeat: matches 164. .307 of consensus"
repeat_region 6305. .6344
repeat_region /note="20 copies 2 mer tt 85% conserved"
repeat_region 7423. .7510
repeat_region /note="44 copies 2 mer at 62% conserved"
repeat_region 8083. .8393
repeat_region /note="AluJ repeat: matches 1. .306 of consensus"
repeat_region 8431. .8734
repeat_region /note="AluJ repeat: matches 1. .301 of consensus"
repeat_region 9829. .10142
repeat_region /note="AluSx repeat: matches 1. .312 of consensus"
misc_feature complement(10324. .10654)
repeat_region /note="match: STS: Em:G55659"
repeat_region 10810. .10977
repeat_region /note="MER53 repeat: matches 2. .189 of consensus"
repeat_region 11378. .11519
repeat_region /note="FLAM_C repeat: matches 1. .136 of consensus"
repeat_region 12045. .12346
repeat_region /note="AluJb repeat: matches 1. .299 of consensus"
repeat_region 12464. .12643
repeat_region /note="AluSp repeat: matches 125. .302 of consensus"
repeat_region 12729. .13015
repeat_region /note="AluSg repeat: matches 3. .298 of consensus"
repeat_region 13285. .13709
repeat_region /note="LTR repeat: matches 172. .691 of consensus"
repeat_region 13777. .14020
repeat_region /note="LTR repeat: matches 1. .256 of consensus"
repeat_region 14196. .14325
repeat_region /note="WLJ1 repeat: matches 3. .136 of consensus"
repeat_region 14739. .14870
repeat_region /note="AluSg/x repeat: matches 1. .132 of consensus"
repeat_region 15479. .15672
repeat_region /note="L2 repeat: matches 2130. .2325 of consensus"
repeat_region 16703. .16838
repeat_region /note="AluSg/x repeat: matches 1. .135 of consensus"
repeat_region 16842. .17132
repeat_region /note="AluSp repeat: matches 1. .299 of consensus"
repeat_region 18017. .18328
repeat_region /note="AluSp repeat: matches 1. .313 of consensus"
repeat_region 18370. .18546
repeat_region /note="AluJo repeat: matches 6. .189 of consensus"
repeat_region 18980. .19084
repeat_region /note="MIR repeat: matches 2. .109 of consensus"
repeat_region 19161. .19450
repeat_region /note="AluSg repeat: matches 2. .299 of consensus"
repeat_region 19464. .19777
repeat_region /note="AluSx repeat: matches 1. .307 of consensus"
repeat_region 19800. .19991
repeat_region /note="MIR repeat: matches 15. .227 of consensus"
repeat_region 19998. .20290
repeat_region /note="AluSg repeat: matches 1. .295 of consensus"
repeat_region 20959. .20994
repeat_region /note="AluSg repeat: matches 134. .169 of consensus"
repeat_region 21390. .21666
repeat_region /note="AluJb repeat: matches 1. .276 of consensus"
repeat_region 21671. .21714
repeat_region /note="Alu repeat: matches 247. .292 of consensus"
repeat_region 23092. .23166
repeat_region /note="FINGER2 repeat: matches 1. .75 of consensus"
repeat_region 23198. .23375
repeat_region /note="MER repeat: matches 1. .177 of consensus"
repeat_region 23496. .23772
repeat_region /note="AluJb repeat: matches 1. .291 of consensus"
repeat_region 24104. .24125
repeat_region /note="11 copies 2 mer at 100% conserved"
repeat_region 24820. .25115
repeat_region /note="AluSg repeat: matches 1. .296 of consensus"
```

```
repeat_region 25506. .25627
repeat_region /note="61 copies 2 mer ct 66% conserved"
repeat_region 26269. .26399
repeat_region /note="L2 repeat: matches 2152. .2294 of consensus"
repeat_region 26466. .26517
repeat_region /note="26 copies 2 mer tt 73% conserved"
repeat_region 26521. .26791
repeat_region /note="AluSg repeat: matches 1. .288 of consensus"
repeat_region 27124. .27416
repeat_region /note="AluSg repeat: matches 1. .293 of consensus"
repeat_region 27646. .27956
repeat_region /note="AluSx repeat: matches 1. .312 of consensus"
repeat_region 28612. .28841
repeat_region /note="L2 repeat: matches 1261. .1463 of consensus"
repeat_region 28842. .29132
repeat_region /note="AluJb repeat: matches 1. .304 of consensus"
repeat_region 29133. .29179
repeat_region /note="L2 repeat: matches 1463. .1513 of consensus"
repeat_region 29232. .29305
repeat_region /note="L2 repeat: matches 2598. .2671 of consensus"
repeat_region 29494. .29619
repeat_region /note="L2 repeat: matches 2617. .2750 of consensus"
repeat_region 29923. .30075
repeat_region /note="MER20 repeat: matches 68. .209 of consensus"
repeat_region 30076. .30200
repeat_region /note="FLAM_A repeat: matches 5. .128 of consensus"
repeat_region 30201. .30255
repeat_region /note="MER20 repeat: matches 12. .68 of consensus"
repeat_region 30360. .30658
repeat_region /note="AluSg repeat: matches 1. .298 of consensus"
repeat_region 32044. .32097
repeat_region /note="27 copies 2 mer tg 100% conserved"
repeat_region 32613. .32907
repeat_region /note="AluSp repeat: matches 3. .300 of consensus"
repeat_region 34314. .34560
repeat_region /note="AluSg repeat: matches 67. .313 of consensus"
repeat_region 35005. .35287
repeat_region /note="AluSg repeat: matches 1. .296 of consensus"
repeat_region 35865. .35914
repeat_region /note="MER53 repeat: matches 1. .50 of consensus"
repeat_region 35962. .36100
repeat_region /note="L2 repeat: matches 1734. .1873 of consensus"
repeat_region 36395. .36428
repeat_region /note="17 copies 2 mer at 88% conserved"
repeat_region 36483. .36510
repeat_region /note="14 copies 2 mer ta 92% conserved"
repeat_region 36925. .37886
repeat_region /note="match: STS: Em:G07212"
repeat_region 38823. .39115
repeat_region /note="AluSg repeat: matches 1. .301 of consensus"
repeat_region 39749. .39772
repeat_region /note="12 copies 2 mer tt 100% conserved"
repeat_region 39794. .39937
repeat_region /note="AluJb repeat: matches 1. .144 of consensus"
repeat_region 39980. .40265
repeat_region /note="AluJo repeat: matches 1. .297 of consensus"
repeat_region 40488. .40782
```

```
Query Match 59.0%; Score 590.2; DB 9; Length 153937;
Best Local Similarity 76.3%; Pred. No. 4.1e-172;
Matches 764; Conservative 0; Mismatches 233; Indels 4; Gaps 3;

Oy 1 GGAGATGATAAACCGTGTGAGTGCCTCAAGTTGTGTGCGACCATGAATGGGAGACTG 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 123176 GGAGATGAGAAACCATGTGGTGCCTCAAGTGTGCGCATGAATGGGAGACTG 123117

Oy 61 GAGGATACATGGATCCCACTACAGGCCCGCAGCTCCTCCAGTATGAGCCATGAGCCAGTT 120
|||| | || | || | || | || | || | || | || | || | || | || | || | || |
Db 123116 GAGGAACCCAGGTGTGCCCAACCATGGCCCGTCCCTCGGTACGAGCCATGAGCCAGCT 123057

Oy 121 GAATCTGAATGTGAAGATGGAATGAAGACCGACGAGAGTTCACACTGACGTCAACCTCAT 180
|||| | || | || | || | || | || | || | || | || | || | || | || | || |
Db 123056 GAGCCTGAGTGCAAAGACGGAGAGAGGCTGACAAAGGTTCAGATGACATCAACCTCAT 122997
```





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 24, 2002, 16:24:42 ; Search time 1812.02 Seconds  
(without alignments)  
16077.036 Million cell updates/sec

Title: US-09-708-724a-3\_COPY\_99000\_100000  
Perfect score: 1001  
Sequence: 1 tggcagcgctgtgtccc.....ttttatccaccatcaactaa 1001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
c 1	1001	100.0	153629	9	AC092325	AC092325 Homo sapi
c 2	990	98.9	191768	9	AC092357	AC092357 Homo sapi
c 3	550.4	55.0	191014	9	AC113268	AC113268 Papio cyn
c 4	494.2	49.4	151049	9	AC018558	AC018558 Homo sapi
c 5	425.6	42.5	2719	9	AK096292	AK096292 Homo sapi
c 6	425.6	42.5	3702	9	BC036456	BC036456 Homo sapi
c 7	284.8	28.5	142978	9	AC097717	AC097717 Homo sapi
c 8	272.2	27.2	167133	2	AC026186	AC026186 Homo sapi
c 9	272.2	27.2	201061	9	AC007353	AC007353 Homo sapi
c 10	129.6	12.9	137496	9	AC005288	AC005288 Homo sapi
c 11	126	12.6	111187	2	AC009283	AC009283 Homo sapi
c 12	126	12.6	161494	2	AC110495	AC110495 Homo sapi
c 13	125	12.5	82453	9	AC004558	AC004558 Homo sapi
c 14	125	12.5	179040	2	AC022732	AC022732 Homo sapi
c 15	124.8	12.5	141594	9	AC068302	AC068302 Homo sapi
c 16	124.8	12.5	190588	2	AC009861	AC009861 Homo sapi
c 17	124.8	12.5	190610	2	AC130419	AC130419 Homo sapi
c 18	123.6	12.3	194995	9	AC018696	AC018696 Homo sapi
c 19	123.4	12.3	171523	9	AC018452	AC018452 Homo sapi
c 20	122.8	12.3	132290	9	AL627402	AL627402 Human DNA
c 21	122.8	12.3	144735	2	AL161458	AL161458 Homo sapi
c 22	122.2	12.2	146443	9	AC007488	AC007488 Homo sapi
c 23	122.2	12.2	163057	2	AC012419	AC012419 Homo sapi
c 24	122	12.2	39645	2	AC110593	AC110593 Homo sapi
c 25	122	12.2	166993	2	AC127522	AC127522 Homo sapi
c 26	122	12.2	194973	9	AC024474	AC024474 Homo sapi
c 27	121.8	12.2	175403	9	AC093268	AC093268 Homo sapi
c 28	121.8	12.2	178182	2	AC114980	AC114980 Homo sapi
c 29	121.8	12.2	178314	2	AC019025	AC019025 Homo sapi
c 30	121.6	12.1	213999	9	AC051619	AC051619 Homo sapi
c 31	121.4	12.1	110000	2	AC021072_1	Continuation (2 of
c 32	120.8	12.1	180490	2	AC069152	AC069152 Homo sapi
c 33	120.8	12.1	192254	2	AC122133	AC122133 Homo sapi
c 34	120.6	12.0	180194	9	AC078850	AC078850 Homo sapi
c 35	120.6	12.0	181531	9	AC007690	AC007690 Homo sapi
c 36	120.6	12.0	183451	9	AC005972	AC005972 Homo sapi
c 37	120.4	12.0	51407	9	AL357556	AL357556 Human DNA
c 38	120.4	12.0	86295	2	AC078810	AC078810 Homo sapi
c 39	120.4	12.0	100000	9	AP000503S2	AP000504 Homo sapi
c 40	120.4	12.0	110000	2	AC125619_4	Continuation (5 of
c 41	120.4	12.0	110000	2	AC027294_2	Continuation (3 of
c 42	120.4	12.0	124156	9	AL596132	AL596132 Human DNA
c 43	120.4	12.0	136493	9	AL662899	AL662899 Human DNA
c 44	120.4	12.0	161413	9	AC119736	AC119736 Homo sapi
c 45	120.4	12.0	163682	9	HSDJ71117	AL132713 Human DNA

ALIGNMENTS

RESULT 1  
AC092325/c  
LOCUS AC092325 153629 bp DNA linear PRI 02-NOV-2001  
DEFINITION Homo sapiens chromosome 16 clone RP11-14K3, complete sequence.  
ACCESSION AC092325 AC011682  
VERSION AC092325.2 GI:16596537  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 153629)  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Human Chromosome 16  
JOURNAL Unpublished

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

2 (bases 1 to 153629)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

3 (bases 1 to 153629)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (02-NOV-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On or before Nov 2, 2001 this sequence version replaced gi:7329393,  
gi:14589514.

COMMENT

Sequence Quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the  
GenBank flat file format but are available as part  
of this entry's ASN.1 file.  
-----  
Sequence Quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the  
GenBank flat file format but are available as part  
of this entry's ASN.1 file.  
-----

FEATURES  
source

Location/Qualifiers  
1..153629  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1b"  
/clone="RP11-14K3"  
BASE COUNT 45079 a 31598 c 31052 g 45900 t

Query Match 100.0%; Score 1001; DB 9; Length 153629;  
Best Local Similarity 100.0%; Pred. No. 4.5e-263;  
Matches 1001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGACGCGCCTGTAGTCCAGCTACTCAGGAGACTGAGCGAGGAGAATCGTTGAACCC 60  
|||||  
Db 130091 TGGCAGCGCGCTGTAGTCCAGCTACTCAGGAGACTGAGCGAGGAGAATCGTTGAACCC 130032  
|||||

QY 61 GGGAGACGGAGGTTGCAGTGAGCCAGAGATCGCGTCACTGCACCTCCAGCCTGGCGAGAC 120  
|||||  
Db 130031 GGGAGACGGAGGTTGCAGTGAGCCAGAGATCGCGTCACTGCACCTCCAGCCTGGCGAGAC 129972  
|||||

QY 121 GTTCCGTTTCAAAAGAAAAAATAATATTAAATAAAGAAATAAAATCCGGCGGTGGCGG 180  
|||||  
Db 129971 GTTCCGTTTCAAAAGAAAAAATAATATTAAATAAAGAAATAAAATCCGGCGGTGGCGG 129912  
|||||

QY 181 TGACATCAGTCTGTGCTGTTAATGCTCGCGCGGTACCGTCCGCGAGTCTCTTCT 240  
|||||  
Db 129911 TGACATCAGTCTGTGCTGTTAATGCTCGCGCGGTACCGTCCGCGAGTCTCTTCT 129852  
|||||

QY 241 GAGGACCCCGCCCGCCACATCTCCGCTTCCCAATAAGAGGTTTCAGGTTTTCGCGTCCGCCGT 300  
|||||  
Db 129851 GAGGACCCCGCCCGCCACATCTCCGCTTCCCAATAAGAGGTTTCAGGTTTTCGCGTCCGCCGT 129792  
|||||

QY 301 GGTGTGTTCTGTGCTGTCACAGGTTGGAACTGGAGATGCCCTCTTCTCTCTCAGGACA 360  
|||||  
Db 129791 GGTGTGTTCTGTGCTGTCACAGGTTGGAACTGGAGATGCCCTCTTCTCTCTCAGGACA 129732  
|||||

QY 361 GRACCATGAGCCTAGCGGCGAGCGCGGTTCGCGAGCTCCCGCTCCGCCAAGCGGCGCT 420  
|||||  
Db 129731 GAACCATGAGCCTAGCGGCGAGCGCGGTTCGCGAGCTCCCGCTCCGCCAAGCGGCGCT 129672  
|||||

QY 421 COTCAGAGCCGTCGTCGTCGCCGCCCTGCGGGAGCTGACCCGAGCGCAGTGCAGAGTTGC 480  
|||||  
Db 129671 COTCAGAGCCGTCGTCGTCGCCGCCCTGCGGGAGCTGACCCGAGCGCAGTGCAGAGTTGC 129612  
|||||

QY 481 TCCTCTTGGCGGTGACACAGGAGGAGCTGGAGCGCGGATCATCTCCAGGAAGCGGCTG 540  
|||||  
Db 129611 TCCTCTTGGCGGTGACACAGGAGGAGCTGGAGCGCGGATCATCTCCAGGAAGCGGCTG 129552  
|||||

QY 541 AGTAGGAACCTGCAGCGCCACATCTCTCTTTACCCGGGGATGTGCAGGATTACCGTGA 600  
|||||  
Db 129551 AGTAGGAACCTGCAGCGCCACATCTCTCTTTACCCGGGGATGTGCAGGATTACCGTGA 129492  
|||||

QY 601 ATCATGACTCGTCATCTCTGCGGAATTTACCAATGGGAAAAATTGGAGTCTAGAAATTTGCC 660  
|||||  
Db 129491 ATCATGACTCGTCATCTCTGCGGAATTTACCAATGGGAAAAATTGGAGTCTAGAAATTTGCC 129432  
|||||

QY 661 TCCATTTTAGCCACCGGTTCCCGAGTAGCTGTTGGGTGATGAAGTCTCCCGAACA 720  
|||||  
Db 129431 TCCATTTTAGCCACCGGTTCCCGAGTAGCTGTTGGGTGATGAAGTCTCCCGAACA 129372  
|||||

QY 721 AATGCGCTGCCATGATAGTTTCTGAAAAAGTAACATGTTTGGTTTCCCGAACAACAATAC 780  
|||||  
Db 129371 AATGCGCTGCCATGATAGTTTCTGAAAAAGTAACATGTTTGGTTTCCCGAACAACAATAC 129312  
|||||

QY 781 AGACTCTGGAGCTTTTAAGCACCTTTATATGTTATAGTTAATGCTTTTAAGTCAGAGTA 840  
|||||  
Db 129311 AGACTCTGGAGCTTTTAAGCACCTTTATATGTTAATGCTTTTAAGTCAGAGTA 129252  
|||||

QY 841 GTTTATCAAGAAAAATTGTAATGATGGAATAAGAGCTCCACAGCATCTAATTTAGAT 900  
|||||  
Db 129251 GTTTATCAAGAAAAATTGTAATGATGGAATAAGAGCTCCACAGCATCTAATTTAGAT 129192  
|||||

QY 901 GTCCAAATTTCTCATCTACTACAAATCATTTCCAGGAAGAAAAAGATAGGACCTTTTGAAA 960  
|||||  
Db 129191 GTCCAAATTTCTCATCTACTACAAATCATTTCCAGGAAGAAAAAGATAGGACCTTTTGAAA 129132  
|||||

QY 961 ATCTGATGATCGGCGCATGTGTTTATCCACCATCACTAA 1001  
|||||  
Db 129131 ATCTGATGATCGGCGCATGTGTTTATCCACCATCACTAA 129091  
|||||

RESULT 2  
AC092357 191768 bp DNA linear PRI 15-AUG-2001  
LOCUS Homo sapiens chromosome 16 clone RP11-332P24, complete sequence.  
DEFINITION AC092357 AC023764  
ACCESSION AC092357.2 GI:15187270  
VERSION HTG.  
KEYWORDS Homo sapiens.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

1 (bases 1 to 191768)  
DOE Joint Genome Institute.  
Sequencing of Human Chromosome 16  
Unpublished

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

2 (bases 1 to 191768)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On or before Aug 15, 2001 this sequence version replaced  
gi:7596817, gi:14589546.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

3 (bases 1 to 191768)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (15-AUG-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On or before Aug 15, 2001 this sequence version replaced  
gi:7596817, gi:14589546.

COMMENT

Sequence Quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than

1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

-----  
Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

FEATURES  
    source  
        1..191768  
            /organism="Homo sapiens"  
            /db\_xref="taxon:9606"  
            /chromosome="16"  
            /clone="RP11-332P24"  
BASE COUNT   57149 a 39592 c 40150 g 54877 t  
ORIGIN

Query Match           98.9%; Score 990; DB 9; Length 191768;  
Best Local Similarity 99.9%; Pred. No. 5e-260;  
Matches 1001; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
  
Qy   1 TGGCAGCGCGCTGTAGTCCAGCTACTCAGGAGACTGAGGAGGAGATCGCTTGAACCC 60  
      |||||  
Db 118183 TGGCAGCGCGCTGTAGTCCAGCTACTCAGGAGACTGAGGAGGAGATCGCTTGAACCC 118242  
      |||||  
  
Qy   61 GGGAGACGAGGTTGCAGTGGCAAGATCGCTCACTGCTCACTCCAGCTCGCGCAGACAGAC 120  
      |||||  
Db 118243 GGGAGACGAGGTTGCAGTGGCAAGATCGCTCACTGCTCACTCCAGCTCGCGCAGACAGAC 118302  
      |||||  
  
Qy   121 GTTCCGTTTCAAAAGAAAAAATATATTAATAAAAAAGATAAATCCGCGCTCGCGG 180  
      |||||  
Db 118303 GTTCCGTTTCAAAAGAAAAAATATATTAATAAAAAAGATAAATCCGCGCTCGCGG 118362  
      |||||  
  
Qy   181 TGACATAGTCTGTGTTAATGCTCGCGCGCTACCGTCTCGCGCAGTCTTCTTCT 240  
      |||||  
Db 118363 TGACATAGTCTGTGTTAATGCTCGCGCGCTACCGTCTCGCGCAGTCTTCTTCT 118422  
      |||||  
  
Qy   241 GAGGA-CGCCCGCCCGCTCTCGCGCTTCCAATAAGAGTTCAAGTTTTCGCGTTCGCGG 299  
      |||||  
Db 118423 GAGGACCGCCCGCCCGCTCTCGCGCTTCCAATAAGAGTTCAAGTTTTCGCGTTCGCGG 118482  
      |||||  
  
Qy   300 TGGTTGCTGTTCTGTGCTGCCACAGTTGGAACTGGAGATGCTTCTTCTCTCAGGAC 359  
      |||||  
Db 118483 TGGTTGCTGTTCTGTGCTGCCACAGTTGGAACTGGAGATGCTTCTTCTCTCAGGAC 118542  
      |||||  
  
Qy   360 AGAACCATGACCTAGCGGAGCGCGCTTCCGAAGCTCCCGTCCGCCAACGGCGCC 419  
      |||||  
Db 118543 AGAACCATGACCTAGCGGAGCGCGCTTCCGAAGCTCCCGTCCGCCAACGGCGCC 118602  
      |||||  
  
Qy   420 TCCTCAGAGCGCTCGTGCCTCGCGCTGCTCGCGGAGCTACCGCGCAGCGAGTTCGAGAGTTG 479  
      |||||  
Db 118603 TCCTCAGAGCGCTCGTGCCTCGCGCTGCTCGCGGAGCTACCGCGCAGCGAGTTCGAGAGTTG 118662  
      |||||  
  
Qy   480 CTCCTGTTGGGGTGACACAGGAGGAGTGGAGCGCGGATCATCTCCAGGAACGGGCT 539  
      |||||  
Db 118663 CTCCTGTTGGGGTGACACAGGAGGAGTGGAGCGCGGATCATCTCCAGGAACGGGCT 118722  
      |||||  
  
Qy   540 GAGTAGGACTGCAGCGCGCCACATCTCTTTTACCGGGGATGTGAGGATACCGTGA 599  
      |||||  
Db 118723 GAGTAGGACTGCAGCGCGCCACATCTCTTTTACCGGGGATGTGAGGATACCGTGA 118782  
      |||||  
  
Qy   600 AATCATGACTCGTCACTCTCGGAATTAACCAATGGGAAAAATGGAGTCTAGAAAAATTATTCG 659  
      |||||  
Db 118783 AATCATGACTCGTCACTCTCGGAATTAACCAATGGGAAAAATGGAGTCTAGAAAAATTATTCG 118842  
      |||||  
  
Qy   660 CTCATTTTAGCCACCGCGTTTCCCGCAGTAGCTGTATTGGGGTGATGAAGTGTCTCCCGAAC 719

|||||  
Db 118843 CTCATTTTAGCCACCGCGTTTCCCGCAGTAGCTGTATTGGGTGATGAAGTCTCCGGAAC 118902  
Qy   720 AATGCGCTGCATGATAGTTTCTGAAAGTAACATGTTTGGTTTCCCAACACACATA 779  
      |||||  
Db 118903 AATGCGCTGCATGATAGTTTCTGAAAGTAACATGTTTGGTTTCCCAACACACATA 118962  
Qy   780 CAGACTCTGGAGCTTTTAAAGCACCTTTATATGTTTATGTTTAAATGCTTTTAAAGTCAGAGT 839  
      |||||  
Db 118963 CAGACTCTGGAGCTTTTAAAGCACCTTTATATGTTTATGTTTAAATGCTTTTAAAGTCAGAGT 119022  
Qy   840 AGTTTATCAAGAGAAATTTGAATGATTGGAATAAGGACTCCACAGCATCTAATTTGTAGA 899  
      |||||  
Db 119023 AGTTTATCAAGAGAAATTTGAATGATTGGAATAAGGACTCCACAGCATCTAATTTGTAGA 119082  
Qy   900 TGTCCAATTTCTCTCATACTACAATCATTTCCAGGAGGAGGAAAGATAGGACCTTTGAAA 959  
      |||||  
Db 119083 TGTCCAATTTCTCTCATACTACAATCATTTCCAGGAGGAGGAAAGATAGGACCTTTGAAA 119142  
Qy   960 AATCTGATGATCGGCCATGTGTTTTTATCCACCATCACTAA 1001  
      |||||  
Db 119143 AATCTGATGATCGGCCATGTGTTTTTATCCACCATCACTAA 119184

RESULT 3  
AC113268/c  
LOCUS           AC113268           191014 bp   DNA   linear   PRI 14-AUG-2002  
DEFINITION      Papio cynocephalus anubis clone rp41-22m16, complete sequence.  
ACCESSION      AC113268  
VERSION        AC113268.8   GI:22095124  
SOURCE        HTG.  
                olive baboon.

ORGANISM       Papio cynocephalus anubis  
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
                Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
                Cercopithecinae; Papio.

REFERENCE       1 (bases 1 to 191014)  
AUTHORS       Zhou,L., Fu,Y., Wu,J., Eichler,E. and Roe,B.A.  
TITLE          Papio anubis BAC Clone rp41-22m16  
JOURNAL        Unpublished

REFERENCE       2 (bases 1 to 191014)  
AUTHORS       Zhou,L., Fu,Y., Wu,J., Eichler,E. and Roe,B.A.  
TITLE          Direct Submission  
JOURNAL        Submitted (28-FEB-2002) Department Of Chemistry And Biochemistry,  
                The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
                OK 73019, USA

REFERENCE       3 (bases 1 to 191014)  
AUTHORS       Zhou,L., Fu,Y., Wu,J., Eichler,E. and Roe,B.A.  
TITLE          Direct Submission  
JOURNAL        Submitted (06-AUG-2002) Department Of Chemistry And Biochemistry,  
                The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
                OK 73019, USA

REFERENCE       4 (bases 1 to 191014)  
AUTHORS       Zhou,L., Fu,Y., Wu,J., Eichler,E. and Roe,B.A.  
TITLE          Direct Submission  
JOURNAL        Submitted (14-AUG-2002) Department Of Chemistry And Biochemistry,  
                The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
                OK 73019, USA

COMMENT        On Aug 4, 2002 this sequence version replaced gi:21104942.  
                ----- Genome Center  
                Center: Department Of Chemistry And Biochemistry  
                The University Of Oklahoma  
                Center code:UOKNOR

FEATURES  
    source  
        1..191014  
            /organism="Papio cynocephalus anubis"  
            /sub\_species="anubis"  
            /db\_xref="taxon:9555"  
            /clone="rp41-22m16"  
            /collection="RP11-332P24"  
BASE COUNT   54882 a 40134 c 39560 g 56438 t

BASE COUNT   54882 a 40134 c 39560 g 56438 t  
ORIGIN

```
Query Match      55.0%; Score 550.4; DB 9; Length 191014;
Best Local Similarity 82.6%; Pred. No. 1.9e-139;
Matches 699; Conservative 0; Mismatches 121; Indels 26; Gaps 5;

QY 1 TGGCAGCGCGCTGTAGTCCAGCTACTCAGAGACTGAGGAGGAGATCGCTTGAACCC 60
Db 76991 TGGCACAAGCCTGTAGTTTCCAGCTAGTCCAGAGACTGAGGAGGAGATCGCTTGAACCC 76932

QY 61 GGGAGCAGGAGTTGACGTGAGCAAGATCGCTCAGTGCACCTCCAGCTGGC---GAC 116
Db 76931 GGGAGTTGGAGTTGACGTGAGCAGCAAGATCGGCGCAGTACACTCAGAGCTGGCAACAGAG 76872

QY 117 AGAGCTTCCGTTTCAAAAGAAAAAAT-----AATATTATTAATAAAAGAAATATCCGGC 171
Db 76871 AGACACTCCGTTTCAAAAAAATAAATAATAATAATAATAATAATAATAATAATAATA 76812

QY 172 GCTGGCGGCTGACATCACTCTCTGCTGCTTAATGCTTCGCGCGGCTACCGTCTCGCGGAG 231
Db 76811 AATAAAATCCCGCGCTGCGCGGTGACGTTGATGTTTCAATGCGGCTTCCGTCAGTGCAG 76752

QY 232 TCTCTTTCTGAGGACCCCGCCCGCCTCCGCTTCCCAATAAGGAGTTCAGGTTTTCG 291
Db 76751 TCCCTTTCTGAGGAC-----CCCGCAGTCTGCGCTTCCCAATAAGGAGCTCAGGTTTTCG 76697

QY 292 GGTCCCGCTGTGCTTCTGCTGTCACAGAGTTGGAAGTGGAGATGCCCTCTCTCTTCT 351
Db 76696 ACTCACTCGGTTGCTTCTGCTGTCGACAGAGTTGGAATGGAGATGCCCTCTCTCTTCT 76637

QY 352 CTCAGGACAGAACCATGAGCTAGGCGGAGCGCGGTTTCGGAAGCTCCCGCTCGGCCAA 411
Db 76636 CTCAGGACAGAACCATGAGCGCGGAGCGGAGAGCTGGCGGCGGAAGCTCCCTCTCAGCCAC 76577

QY 412 CGGGCGCTCTCAGAGCGCTGCGTGCGCGCCCTGCGGGGAGCTGACCGCGAGCGACGTG 471
Db 76576 TGGGACCGCTCAGAGCCCGACCGTGTCCGCTGCGCGGAGCTGACCT-AGCGCTGTG 76518

QY 472 CAGAGTTCTCTGTTGGCGGTGACCGAGGAGGAGCTGGAGCGGCGGATCATCTCCAGGA 531
Db 76517 CAGAGTTCTCTGCTGCTGACCGAGGAGGAGCTGGAGTGGCGGATCATCTCCAGGA 76458

QY 532 AGCGGCTGAGTAGGAATGAGCGCGGACACATCTCTCTTTACCGGGGATGTGCAGGAT 591
Db 76457 AGAGAGCTGAGTAGGAATGAGCGCGGACACGCTCTCTTTCCCGGGGATGTGCAGGAT 76398

QY 592 TACCGTGAATATCAGCTGCTCATCTCGGAATTACCAATGGAAATTTGGAGTCTAGAA 651
Db 76397 TACTGTGAATATGAGCTGCTCATCTCGGAATTACCAATGGAAATTTGGAGTCTAGAA 76338

QY 652 ATTATGCTCCATTTTAGCCCCACCGTTCCCGAGTAGCTGTATTGGGGTGTGAAGTGC 711
Db 76337 ATTATGAGCCATTTTAGCCCCACCGTTCCCGAATAGCAGTATTGGGTGTGAAGTGC 76278

QY 712 TCC-----GGACAATGCGGTGCGCATGATAGTTTTCTGAAAGTAACATGTTT 760
Db 76277 TCCCAAGCACATTTTCACAAAATTCAGCTGCGCATGATAGTTTTCTGAAAGTAACATGTTT 76218

QY 761 GGTTCCTCCAGAACACAATACAGACTCTGAGCTTTTAAGCACTTTATATGTTATTAGTT 820
Db 76217 GGTTCCTCTGAAGGCAATACAGATCTGGAGCTTTTAAGCACTTTATATGTTATTAGTT 76158

QY 821 AATGCT 826
Db 76157 TGTGCT 76152

RESULT 4
AC018558
LOCUS AC018558 151049 bp DNA linear PRI 02-NOV-2001
DEFINITION Homo sapiens chromosome 16 clone RP11-80F22, complete sequence.
ACCESSION AC018558
VERSION AC018558.5 GI:16596530
KEYWORDS HTG.
```

```
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Mammalia; Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 151049)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 151049)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (14-DEC-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 151049)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (02-NOV-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Nov 2, 2001 this sequence version replaced gi:9795566.
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
Genbank flat file format but are available as part
of this entry's ASN.1 file.
-----
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
Genbank flat file format but are available as part
of this entry's ASN.1 file.
-----
FEATURES             Location/Qualifiers
     source            1..151049
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="16"
                        /clone="RP11-80F22"
BASE COUNT 46827 a 30783 c 30749 g 42690 t
ORIGIN
Query Match      49.4%; Score 494.2; DB 9; Length 151049;
Best Local Similarity 76.1%; Pred. No. 4.7e-124;
Matches 689; Conservative 0; Mismatches 178; Indels 38; Gaps 5;

QY 132 AAAGAAAAAATAATATTATAAAAGAAATAAATCCGGCGCTCGCGGTGACATCAGTC 191
Db 40966 AATGGAATGAGATTAAACACACAAATAATGAAATCCAGCGCTCGCGCTGACATCGCC 41025

QY 192 TCTGCTGTTAATGCTCGCGCGGTACCGTCTCT-----GCGCA 230
Db 41026 TCTGAAGTGTTCGCCGCACTGGCTTCTGGCTGGCGGTGCGCGACAGTTGGCGCTC 41085

QY 231 GTCTCTTTCTGAGGACCCCCCCCCCACTCTCCGCGCTTCCAATAAGAGAGTTTCAGGTTTTG 290
Db 41086 CTGCAGCAGTCCCAACCTCAGGACTCCCTACCTCCACCTCCAAACAGCTCGGTTCTTTG 41145

QY 291 CGTCCGCGTGTGCTGCTCTCTGCTGCCACAGAGTTGGAAGTGGAGATGCTCTTCCTTC 350
Db 41146 CGTCCGCGCGGCTGCTGCTCTGCTCTCACGTTTGAATTTGGAGATGCTCTTCCTTCG 41205

QY 351 TCTCAGGACAGAACCATGAGCTTAGCGGACGCGGCTTCGCGAAGTTCGCCCTCCGCA 410
Db 41206 TCTCAGGACAGAACCATGAGAACACGACGCGGCGGTGCAGATGCTCCCAATCCGTCG 41265

QY 411 ACGGCGCGCTCTCTCAGAGCC--GTCCGTGCCCGCGCTGCGGAGCTGACCCGACGCGAG 469
```





```

Db 634 TTTTGTGAAAAGTAACATGTTTGGTGGCCCCAGAACACAACTACTGACTTTGGAGCTTTTAA 693
Qy GCACCTTTATATGTTATAGTAAATGCTTTTA-----AGTCAGAGTAGTCTTATCAAAAGGA 853
Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
694 GCACCTTATATGTTATAGTAAATGCTTTTAATTTAGTCAGATAGTATTTATCAAAAGAA 753
Qy 854 AAATTTGAATGATTGGAATAAGGACTCCACAGCACTAAATTAATTTAGATGTCCAATCTCTT 913
Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
754 AAGTTGAATGTTTGAATAAGGACTCCATAGCATCTAACTAGTAGA--TCCAGCTCTTCT 811
Qy 914 CATACTACAATCATTTCCAGAGAGGAAAGATAGGACCTTTGAAAATCTGATGGATCG 973
Db ||||||| || || || || || || || || || || || || || || || || || || || || || ||
812 CATACTAGGAATGTTGCCAGGAGAAAGTAGGAGCCTGTGAAAATCTGATGAGTCT 871
Qy 974 GCCATGTGTTTTATCCACCACACTAA 1001
Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
872 GCCATGATTTTTATCCACCACACTAA 899

RESULT 6
BC036456 3702 bp mRNA linear PRI 12-AUG-2002
LOCUS Homo sapiens, clone MGC:33409 IMAGE:4825616, mRNA, complete cds.
DEFINITION BC036456
ACCESSION BC036456.1 GI:22209023
VERSION human.
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3702)
Direct Submission
Submitted (09-AUG-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
contact: amadan@systemsbio.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 46 Row: n Column: 15
This clone was selected for full length sequencing because it
passed the following selection criteria: GenomeScan gene
prediction.

FEATURES
source
1..3702
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:33409 IMAGE:4825616"
/tissue_type="Testis"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="vector: pBluescript"
487..1296
/codon_start=1
/product="Unknown (protein for MGC:33409)"
/protein_id="AAH36456.1"
/db_xref="GI:22209024"
/translation="MTRPENYQWENWLENVATILAHFPNSYIWIWIKSRMLHKF
SCYDNFVKSNNFGAPEHDFGAKHLYMLLVNFAFNLSQNSLKSUNWNKOSIASN
CRSSPSHTTNGCQGEKVRTEKSDSAMSFYPSLNDASFTLLIGFSKGCYVLNQLLFE

REFERENCE
AUTHORS
TITLE
JOURNAL

REMARK
COMMENT

```

```

LKEAKDKNIDAFIKSIRTWYLDGHSNGSNWTWYPEVLKFEFAQTGIIIVHTVTPY
QVDRPNRSWIGKEHKFQILGDMQVTSQHFTKEAPSIENHFRVHFV"
BASE COUNT 1161 a 655 c 728 g 1158 t
ORIGIN
Query Match 42.5%; Score 425.6; DB 9; Length 3702;
Best Local Similarity 79.1%; Pred. No. 1.3e-105;
Matches 592; Conservative 0; Mismatches 124; Indels 32; Gaps 6;

Qy 284 GGTTTTCCGGTCGCCGTGTTGCTCTCTGCCACAGGTTCGAACTGGAGATGCCTC 343
Db || || || || || || || || || || || || || || || || || || || || || ||
153 GCTCCTCGGTGCGCGCGTGTGCTGCTGCCGAGCTCGGAATCGGAACGGCTC 212
Qy 344 TTCCTTCTCTCAGACAGAACCATG-AGCCTAGCGGAGCGGCCGTTTCGGGAAGCTCCCC 402
Db || || || || || || || || || || || || || || || || || || || || || ||
213 GTCTGTCTCTCAGCCAGAACCATGAAACCGGGCGGCGGCGCGGCTCTCCT 272
Qy 403 CTCCGCCAACGGCGCTCTCAGAGCCGTCC--GTGCCGCCCTGCGGGAGAGTGACCC 460
Db || || || || || || || || || || || || || || || || || || || || || ||
273 CTGGCGGAGTGGCGCGCTCAGTGCCTGCAGCTTTCCACCGTGCCTGGAGCCGATCC 332
Qy 461 GCAGCGAGTCGACAGTTGCTCTGTTGGCGGTCAACAGGAGGAGGACTGGAGCGCGGAT 520
Db || || || || || || || || || || || || || || || || || || || || || ||
333 GCAGCGCAGCAACGAATGCTCTGTTGGCGGCGCGCGGAGGAGGACTGGAGCGCGAGGA 392
Qy 521 CATCTCCAGGAAGCGGCTGAGTAGGAAGTGCAGCCGC-----CACATCCTCTCTC 569
Db || || || || || || || || || || || || || || || || || || || || || ||
393 CCTCCCGGGACCCAGGAGGAGGAGCCGACGCCGCCCGCCAGCATCAGCTCTCTA 452
Qy 570 TTTACCGGGGATGTGAGGATTAACGTGAATCATGACTGTCATCTCTCGGAATTACCA 629
Db || || || || || || || || || || || || || || || || || || || || || ||
453 TTTCCCTGGGATGTGAGAAATTAACATGAATTAATGACTGCTCATCTGAGAAATTAATCA 512
Qy 630 ATGGGAAATTTGGAGTCTAGAAATTAATGCTCCATTTAGCCACCGCTTCCCGAGTAG 689
Db || || || || || || || || || || || || || || || || || || || || || ||
513 ATGGGAAACTGGAGTCTAGAAATTTAGTACCATTTTAGCCACCGGTTTCCCAATAG 572
Qy 690 CTGTATTGGGTGATGAAGTCTCCGGAA-----CAAAATGGCTGCCCATGATAG 738
Db || || || || || || || || || || || || || || || || || || || || || ||
573 TTATTATTGGGTGATAAAATGTTCCCGAATGCAATTTGCACAAATTCAGCTCTATGACAA 632
Qy 739 TTTTCTGAAAGTAACATGTTGTTGTTCCAGAACACAAATACAGACTCTGGAGCTTTTAA 798
Db 633 TTTTGTGAAAGTAACATGTTGTTGTTCCCGCCAGAACAACTACTGACTTTGGAGCTTTAA 692
Qy 799 GCACCTTTATATGTTATAGTAAATGCTTTTA-----AGTCAGAGTAGTTATCAAAAGGA 853
Db || || || || || || || || || || || || || || || || || || || || || ||
693 GCACCTTTATATGTTATAGTAAATGCTTTTAATTTAAGTCAGAAATAGTTATCAAGAA 752
Qy 854 AAATTTGAATGATTGGAATAAGGACTCCACAGCATCTAAATTTAGATGTCCAATCTCTTCT 913
Db || || || || || || || || || || || || || || || || || || || || || ||
753 AAGTTGAATGTTGGAATAAGGACTCCATAGCATCTAACTGTAGA--TCCAGTCTTCT 810
Qy 914 CATACTACAATCATTTCCAGGAGGAAAGATAGGACCTTTGAAAATCTGATGGATCG 973
Db || || || || || || || || || || || || || || || || || || || || || ||
811 CATACTAGGAATGTTGCCAGGAGAAAGTAGGAGCCTGTGAAAATCTGATGAGTCT 870
Qy 974 GCCATGTGTTTTATCCACCACACTAA 1001
Db || || || || || || || || || || || || || || || || || || || || || ||
871 GCCATGATTTTTATCCACCACACTAA 898

RESULT 7
AC097717 AC097717 142978 bp DNA linear PRI 21-FEB-2002
LOCUS Homo sapiens BAC clone RP11-650N19 from 2, complete sequence.
DEFINITION AC097717 AC037452
ACCESSION AC097717.3 GI:18072229
VERSION HTG.
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```







```
Db 96480 CACCACCCCTCCGCTCCCAATAGCTCGGGTTTTTTTTTTTGGGTCGTGCGGCTGCTGCT 96421
QY 311 CTTGTGTCACAGGTTGGAAGTGGAGATGCCCTCTCTCTCTCAGGACAGAACCATGAG 370
||| ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 96420 CCTACTGCGGCACACTTGGAAATCCAAACGCTTCGTGGTCTCTCAGGACAGAACCATGAA 96361
QY 371 CTTAGGCGGACGCGCGGTCGCGGAAGCTCCCTCCGCAACAGGCGCTCTCTCAGAGCC 430
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 96360 CCGAGC-GCAGCTGCGGTGCGTGACACCGCC-----AGGCACCGCCTGCGGAGCC 96312
QY 431 GTCCGTGCGCCGCTCCGCGGAGCTGACCCGCGAGCGAGTCAGAGTTCCTCTGCTGGC 490
||| ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 96311 TGACGTGTCTCTCTTGGCTGAGCCACCCGCGAGCGAGCTGCTCTCTCTGCTAGC 96252
QY 491 GGTGACACGAGGAGGACTGAGCGCGGATCATCTCCAGAAAGCGGCTGAGTAGGAATC 550
||| ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 96251 GGTGCGCGGAGGAGTGGAGCGCGGACACCTTCGGGGAGCGTGCAAGGTGGAACC 96192
QY 551 CGAGCGCGCACATCTCTCTTTACCC-GGGGATGTCAGGATTTACCGTGAATCATGACT 609
||| ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 96191 CGAGCGCGCGCACCTCTCTCTCCCTGGGATGTGCAAAATTACCATGAATTAAGCT 96132
QY 610 CTTCTATCTGCAATTAACCAATGGGAAATTTGAGTCTAGAAATTTATGCCCTCCATTTA 669
||| ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 96131 TGTCTATCTGCGAGTTACAGTTGGAATTTGAGTCTAGAAATTTGTTGCCCCCATTTTA 96072
QY 670 GCCACCGGTTCCCAAGTACGTGTTGGGTGATGAAGTCTCGGGAAC-----719
||| ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 96071 GCCACCGGTTCCCAATGGCTATATTGGGTGATAAAGTGTCCGACGCAATTGAC 96012
QY 720 -AAATGCGTGCATGATAGTTTTCTGAAAGTACATGTTGGTTTCCCAAGACAAAT 778
||| ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 96011 AAATTCAGCTGCGGTAGTATTTCTGAAAGTAAACATATTGGTTGCCCAAGACAAAT 95952
QY 779 ACAGACTCTGGAGCTTTTAAGCACCTTTATATATTAGTTA 821
||| ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 95951 ACTACATATTAATTTTATAACATTTATATATATTATATA 95909

RESULT 9
AC007353/c
LOCUS Homo sapiens chromosome 16 clone RP11-488I20, complete sequence.
ACCESSION AC007353
VERSION AC007353.4 GI:14277247
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,
TITLE Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,
JOURNAL Bryant,J., Tesker,J., Meincke,L., Longmire,J., White,S., Tatam,O.,
AUTHORS Campbell,C., Fawcett,J., Maltbie,M., Bussod,M., Sutherland,R.,
McMurry,K., Han,C. and Deaven,L.
Unpublished
2 (bases 1 to 201061)
1 (bases 1 to 201061)
Sequencing of Human Chromosome 16
REFERENCE
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
2 (bases 1 to 201061)
1 (bases 1 to 201061)
Direct Submission
Submitted (22-APR-1999) Center for Human Genome Studies, DOE Joint
Genome Institute, Los Alamos National Laboratory, MS M888, Los
Alamos, NM 87545, USA
3 (bases 1 to 201061)
DOE Joint Genome Institute.
Direct Submission
Submitted (01-JUN-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jun 1, 2001 this sequence version replaced gi:9795555.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
```

Web site: <http://www.jgi.doe.gov>

#### Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

#### FEATURES

Location/Qualifiers

1..201061  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="RP11-488I20"

BASE COUNT 57062 a 41994 c 41012 g 60993 t  
ORIGIN

Query Match 27.2%; Score 272.2; DB 9; Length 201061;  
Best Local Similarity 73.1%; Pred. No. 4e-63;  
Matches 426; Conservative 0; Mismatches 133; Indels 24; Gaps 5;

QY 252 CCCCACCTCTCCGCTTCCAATA-AGGAGTTTCAGGTTTTTGGCGTCCGCTGGTGTGCTGTT 310  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 45886 CACCACCTCCGCTCCCAATAGCTCGGGTTTTTTTTTGGCGTCTGTCGCGCTGCTGCT 45827

QY 311 CTTGTGTCACAGTGTGGAACCTGAGATGCCCTTCCTCTCTCAGGACAGAACCATGAG 370  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 45826 CTTACTCGGCAACTTGAATCCAAACGCTTCGTGGTCTCTCAGGACAGAACCATGAA 45767

QY 371 CCTACGCGCAGCGGTCGCGAAGCTCCCTCCGCAACGCGGCTCTCTCAGAGCC 430  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 45766 CCGAGC-GCAGCTGCGGTGCGTGACACCGCC-----AGGCACCGCTCGGAGCC 45718

QY 431 GTCCGTGCGCCGCTCCGCGGAGCTGACCCGCGAGCGAGTCAGAGTTCCTCTCTGTCG 490  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 45717 TGACGTGTCTCTCTTGGCTGAGCCACCCGCGAGCGAGCTGCTCTCTGCTAGC 45658

QY 491 GGTGACACGAGGAGGACTGAGCGCGGATCATCTCCAGGACGCGGCTGAGTAGGAATC 550  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 45657 GGTCTCCGCGGAGGAGCTGAGCGCGGACCTTCGCGGAGCGTGAAGGTGGAACC 45598

QY 551 CGAGCGCGCACATCTCTCTTTACCC-GGGGATGTGCAAGATTACCGTGAATCATGACT 609  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 45597 GCAGCGCGCGCACCTCTCTCTCCCTGGGATGTGCAAAATTACCATGAATTAAGCT 45538

QY 610 CTTCTATCTGCAATTAACCAATGGGAAATTTGAGTCTAGAAATTTATGCCCTCCATTTA 669  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 45537 TGTCTATCTCGAGTTACCAAGTTGGAATTTGAGTCTAGAAATTTGTTGCCCCCATTTTA 45478

QY 670 GCCACCGGTTCCCGAGTACGTGTTGGGTGATGAAGTCTCGGAAC-----719  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 45477 GCCACCGGTTCCCGAATGGCTATATTTGGGTGATAAAGTGTTCGCGAAGCATTTGAC 45418

QY 720 -AAATGCGTGCATGATAGTTTTCTGAAAGTAAACATGTTGGTTTCCCAAGACAAAT 778  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 45417 AAATTCAGCTGCGGTAGTATTTCTGAAAGTAAACATATTGGTTGCCCAAGACAAAT 45358

QY 779 ACAGACTCTGGAGCTTTTAAGCACCTTTATATATTAGTTA 821  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 45357 ACTACATATTAATTTTATAACATTTATATATATTATATA 45315

#### RESULT 10

AC005288/c  
LOCUS Homo sapiens chromosome 17, clone hCIT.131\_K.11, complete sequence.  
DEFINITION AC005288  
ACCESSION AC005288  
VERSION AC005288.1 GI:3492893  
KEYWORDS HTG.

SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 137496)		
JOURNAL	Birren,B., Linton,L., Nusbaum,C. and Lander,E.		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 137496)		
	Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,		/rpt_family="AT-rich"
	Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boatn,C.,		9610. .9633
	Boutwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E.,		/rpt_family="AT-rich"
	Devon,K., Dewar,K., Donelan,L., Etemadi,S., Ferreira,P.,		9640. .9939
	FitzHugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S.,		/rpt_family="AluSg"
	Gensheimer,S., Gerstegery,K., Gilmartin,T., Grant,G., Hagos,B.,		complement(10233. .10312)
	Harris,K., Horton,L., Howland,J.C., Hui,L., Jacotot,L., Kann,L.,		/rpt_family="MIR"
	Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,		complement(10313. .10626)
	Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,		/rpt_family="AluSg"
	Nachman,A., Nahf,R., Naylor,J., Niloff,M., O'Connor,T., Pavlin,B.,		complement(10627. .10779)
	Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R.,		/rpt_family="MIR"
	Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,		complement(10869. .10942)
	Strickland,C., Subramanian,A., Torruella-Miller,I., Vassiliev,H.,		/rpt_family="AluJ/FLAM"
	Vo,A., Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and		11407. .11546
	Zody,M.		/rpt_family="AluSg/x"
TITLE	Direct Submission		complement(11636. .11937)
JOURNAL	Submitted (16-JUL-1998) Whitehead Institute/MIT Center for Genome		/rpt_family="AluSg"
REFERENCE	Research, 320 Charles Street, Cambridge, MA 02141, USA		complement(11993. .12298)
AUTHORS	3 (bases 1 to 137496)		/rpt_family="AluSg"
	Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,		complement(13171. .13469)
	Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J.,		/rpt_family="AluSx"
	Boutwell,C., Brown,A., Castle,A., Cerny,J., Colangelo,M.,		13754. .14055
	Collins,S., Collymore,A., Cooke,P., Corliss,D., Depayre,E.,		/rpt_family="AluSx"
	Devon,K., Dewar,K., Donelan,L., Ferreira,P., FitzHugh,W.,		14058. .14333
	Forrest,C., Funke,R., Gage,D., Gardyna,S., Gerstegery,K., Grant,G.,		/rpt_family="AluSx"
	Hagos,B., Heaford,A., Herena,B., Horton,L., Howland,J.C.,		15277. .15305
	Jacotot,L., Jones,C., Kann,L., Karatas,A., Lehoczky,J.,		/rpt_family="AT-rich"
	Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,		complement(15306. .15604)
	Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,		/rpt_family="AluSg"
	Nahf,R., Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P.,		complement(15607. .15881)
	Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A., Severy,P.,		/rpt_family="AluSg"
	Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,		complement(16147. .16454)
	Subramanian,A., Tesfaye,S., Tichovolisky,N., Torruella-Miller,I.,		/rpt_family="AluSg"
	Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,		complement(16456. .16750)
	Ye,W.J., Zhao,J. and Zody,M.		/rpt_family="AluSg"
TITLE	Direct Submission		17102. .17164
JOURNAL	Submitted (29-AUG-1998) Whitehead Institute/MIT Center for Genome		/rpt_family="MIR"
REFERENCE	Research, 320 Charles Street, Cambridge, MA 02141, USA		complement(17326. .17472)
COMMENT	On Aug 29, 1998 this sequence version replaced gi:3449085.		/rpt_family="FLAM_C"
	All repeats were identified using RepeatMasker: Smit, A.F.A. &		complement(17561. .17845)
	Green, P. (1996-1997)		/rpt_family="AluSg1"
	<a href="http://ftp.genome.washington.edu/RM/RepeatMasker.html">http://ftp.genome.washington.edu/RM/RepeatMasker.html</a> .		complement(17877. .18033)
FEATURES	Location/Qualifiers		/rpt_family="AluSg"
source	1. .137496		complement(18034. .18337)
	/organism="Homo sapiens"		/rpt_family="AluSx"
	/db_xref="taxon:9606"		complement(18338. .18456)
	/chromosome="17"		/rpt_family="AluSg"
	/map="17"		complement(18770. .18794)
	/clone="hCIT_131_K_11"		/rpt_family="(TAAA)n"
	1-1941"		complement(18795. .19075)
	/clone_lib="Research Genetics/Cal Tech CIB978SK-B (plates		complement(19483. .19791)
	1-1941)"		/rpt_family="AluSg"
repeat_region	1391. .1437		19832. .20136
repeat_region	/rpt_family="purine-rich"		/rpt_family="AluSg"
repeat_region	complement(4525. .4815)		complement(20142. .20437)
repeat_region	/rpt_family="AluSg"		/rpt_family="AluSg"
repeat_region	5059. .5187		complement(20670. .20734)
repeat_region	5189. .5442		/rpt_family="MER47A"
repeat_region	/rpt_family="AluJo/FLAM"		20735. .21033
repeat_region	5189. .5442		/rpt_family="AluSg"
repeat_region	6296. .6354		21038. .21334
repeat_region	/rpt_family="(CGG)n"		complement(21335. .21614)
repeat_region	7179. .7209		/rpt_family="MER47A"
repeat_region	/rpt_family="GC-rich"		21961. .21982
repeat_region	9221. .9265		/rpt_family="AT-rich"
			complement(21983. .22265)
			/rpt_family="AluSp"
			complement(22281. .22303)
			/rpt_family="(TAAA)n"
			complement(22304. .22548)
			/rpt_family="AluSg1"

```

repeat_region complement(22567..22862)
repeat_region /rpt_family="AluSc"
repeat_region complement(23289..23350)
repeat_region /rpt_family="AluJ/FRAM"
repeat_region complement(23385..23541)
repeat_region /rpt_family="FRAM"
repeat_region complement(23636..23937)
repeat_region /rpt_family="AluSx"
repeat_region 24024..24067
repeat_region /rpt_family="AT_rich"
repeat_region complement(24068..24366)
repeat_region /rpt_family="AluSx"
repeat_region complement(24369..24660)
repeat_region /rpt_family="AluJo"
repeat_region 24661..24711
repeat_region /rpt_family="AT_rich"
repeat_region 24723..25021
repeat_region /rpt_family="AluSg"
repeat_region complement(25140..25272)
repeat_region /rpt_family="FLAM_A"
repeat_region complement(25427..25737)
repeat_region /rpt_family="AluSc"
repeat_region complement(25774..26059)
repeat_region /rpt_family="AluJo"
repeat_region complement(26150..26450)
repeat_region /rpt_family="AluJ"
repeat_region 26454..26487
repeat_region /rpt_family="AT_rich"
repeat_region complement(26498..26797)
repeat_region /rpt_family="AluSg"
repeat_region complement(26798..27092)
repeat_region /rpt_family="AluJb"
repeat_region complement(27113..27414)
repeat_region /rpt_family="AluJo"
repeat_region complement(27439..27734)
repeat_region /rpt_family="AluSg"
repeat_region complement(27812..28598)
repeat_region /rpt_family="L2"
repeat_region 28596
repeat_region /note="Probably A"
repeat_region complement(28623..28805)
repeat_region /rpt_family="AluJo"
repeat_region complement(28806..29111)
repeat_region /rpt_family="AluJo"
repeat_region complement(29112..29221)
repeat_region /rpt_family="AluJo"
repeat_region complement(29228..29364)
repeat_region /rpt_family="L1M4"
repeat_region complement(29375..29666)

Query Match 12.9% Score 129.6; DB 9; Length 137496;
Best Local Similarity 81.5% Pred. No. 4.8e-24;
Matches 150; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

Oy 1 TGGCAGCGGCTGTAGTCCAGCTACTCAGGAGACTGAGGAGGAGAGATCGCTTGAACCC 60
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 22717 TGGCGTGTGCTGTAGTCCAGCTACTCAGGAGGCTGAGGAGGAGAGATCGCTTGAACCC 22658

Oy 61 GGGAGACGAGGTTCAGTGTAGCCCAAGATCGGTCACCTGCACCTCCAGCTGGCGACACAC 120
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 22657 GGGAGCAGAGGTTGTCAGTGTAGCCCAAGATCGGCTGTCACCTGCACCTGACACACAG 22598

Oy 121 GTTCGGTTTCAAGAGAAAAAATATATATATAAAAGAAATAAAATCCGGCGCTCGCGGG 180
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 22597 ATTGCTCTCAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATA 22538

Oy 181 TGAC 184
|||
Db 22537 TGGC 22534

RESULT 11
AC009283/c

```

---

LOCUS AC009283 111187 bp DNA linear HTG 19-AUG-2002  
DEFINITION Homo sapiens clone RP11-390P24, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 2  
unordered pieces.  
AC009283  
VERSION AC009283.3 GI:22297433  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_FULLTOP; HTGS\_ACTIVEFIN.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
1 (bases 1 to 111187)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, J., Brown, A., Castle, A., Cerny, J., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearellano, K., Depayre, E., FitzHugh, W., Forrest, C., Donelan, L., Doyle, M., Ferreira, P., FitzHugh, W., Gilbert, D., Grant, G., Funke, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Jones, C., Kann, L., Karatas, A., Lebecky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J., Moll, M., Morris, W., Morrow, J., Mychaleckyj, J., Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B., Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P., Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A., Tesfaye, S., Toruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.  
Direct Submission  
Submitted (12-AUG-1999), Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 111187)  
Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouhgalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission

Submitted (19-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Aug 19, 2002 this sequence version replaced gi:17386471.  
All repeats were identified using RepeatMasker:  
Smith, A. F. A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence.submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L831  
Center clone name: 390\_P\_24  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 2 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.





Df 105814 GGAGCGGAGGTTCAGTGCATCCACCTGCCACCATTGCATCCACTGGCAGCAA 105873

QY 121 GTTCGGTTCAAAAGAAAAATAAATAATTAATAAAAAGAATAAAAATCCGGCGCT 174  
|| || || || | ||| || || || || || || || ||  
Df 105874 ACTCTGCTCAAAATTAATAAAAAATAAATAAATAAATAAATAAATAAATGTTGCT 105927

RESULT 13  
LOCUS AC004558  
DEFINITION Homo sapiens chromosome 19, DNA linear PRI 15-APR-1998 complete sequence.  
ACCESSION AC004558  
VERSION AC004558.1 GI:30471130  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
  
REFERENCE Lamerdin,J.E., McCreedy,P.M., SKowronski,E., Adamson,A.W., Burkhardt-Schultz,K., Gordon,L., Kyle,A., Ramirez,M., Stilwagen,S., Phan,H., Velasco,N., Gurnes,J.J., Danganan,L., Poundstone,P., Christensen,M., Georgescu,A., Avila,J., Liu,S., Attix,C., Andreise,T., Frankheim,M., Amico-Keller,G., Coefield,J., Duarte,S., Lucas,S., Bruce,R., Thomas,P., Quan,G., Kronmiller,B., Arellano,A., Montgomery,M., Ow,D., Nolan,M., Trong,S., Kobayashi,A., Olsen,A.O. and Carraro,A.V.  
Sequence analysis of a 2.5 Mb region in lqql3.2 containing a clustered CEA/PSG gene family Unpublished  
2 (bases 1 to 82453)  
Lamerdin,J.E.  
Direct Submission  
Submitted [15-APR-1998] Joint Genome Institute, Lawrence Livermore National Laboratory, 7000 East Ave., Livermore, CA 94551, USA  
Map and sequence oriented from q centromere to telomere. Accession comprised of sequence from cosmid F20014 from bases 1 to 38,269 and cosmid P8998 from bases 38,270 to 82,453. No sequence errors were detected in overlapping region. Currently there is a small sequence gap between cosmid F20014 and F9933 to the left. Cosmid F9998 overlaps cosmid F24083 to the right by approx. 4 kb.  
Additional map and sequence information may be obtained at:  
<http://www.bio.llnl.gov/genome/genome.html>.  
  
FEATURES source  
Location/Qualifiers  
1..82453      /organism="Homo sapiens"  
                /db\_xref=taxon:9606"  
                /chromosome="19"  
                /map="BCKDHA-D19S217"  
                /clon="F20014-F8998"  
                /cell\_line="UV5HL9-5B"  
                /clone\_lib="LL19NC02 F chromosome 19-specific cosmid library"  
                /note="Cosmid library constructed at LLNL from flow-sorted chromosomes from hybrid UV5HL9-5B, which carries chromosome 19 as its only human chromosome"  
complement(751..1049)  
                /rpt\_family="Aluy"  
                1050..1178  
                /rpt\_family="MIR"  
complement(1959..2141)  
                /rpt\_family="MIR"  
2253..2443  
                /rpt\_family="MLTID"  
complement(2867..3390)  
                /rpt\_family="MER57\_internal"  
complement(3559..3878)  
                /rpt\_family="MER65\_internal"  
3881..4168  
                /rpt\_family="Aluy"  
complement(4173..4665)  
                /rpt\_family="MER65\_internal"



http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L5546

Center clone name: 399.O.18

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 164115 bases at least Q40

Consensus quality: 170430 bases at least Q30

Consensus quality: 173222 bases at least Q20

Insert size: 188000; agarose-fp

Insert size: 175940; sum-of-contigs

Quality coverage: 5.2 in Q20 bases; agarose-fp

Quality coverage: 5.6 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 32 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 1289: contig of 1289 bp in length  
\* 1290 1389: gap of 100 bp  
\* 1390 2987: contig of 1598 bp in length  
\* 2988 3087: gap of 100 bp  
\* 3088 4215: contig of 1128 bp in length  
\* 4216 4315: gap of 100 bp  
\* 4316 5772: contig of 1457 bp in length  
\* 5773 5872: gap of 100 bp  
\* 5873 8065: contig of 2193 bp in length  
\* 8066 8165: gap of 100 bp  
\* 8166 10389: contig of 2224 bp in length  
\* 10390 10489: gap of 100 bp  
\* 10490 12485: contig of 1996 bp in length  
\* 12486 12585: gap of 100 bp  
\* 12586 15882: contig of 3297 bp in length  
\* 15883 15982: gap of 100 bp  
\* 15983 19042: contig of 3060 bp in length  
\* 19043 19142: gap of 100 bp  
\* 19143 21822: contig of 2680 bp in length  
\* 21823 21922: gap of 100 bp  
\* 21923 24967: contig of 3045 bp in length  
\* 24968 25067: gap of 100 bp  
\* 25068 29273: contig of 4206 bp in length  
\* 29274 29373: gap of 100 bp  
\* 29374 32046: contig of 2673 bp in length  
\* 32047 32146: gap of 100 bp  
\* 32147 35253: contig of 3107 bp in length  
\* 35254 35353: gap of 100 bp  
\* 35354 39343: contig of 3990 bp in length  
\* 39344 39443: gap of 100 bp  
\* 39444 43061: contig of 3618 bp in length  
\* 43062 43161: gap of 100 bp  
\* 43162 46849: contig of 3688 bp in length  
\* 46850 46949: gap of 100 bp  
\* 46950 48108: contig of 1159 bp in length  
\* 48109 48208: gap of 100 bp  
\* 48209 52614: contig of 4406 bp in length  
\* 52615 52714: gap of 100 bp  
\* 52715 58023: contig of 5309 bp in length  
\* 58024 58123: gap of 100 bp  
\* 58124 62900: contig of 4777 bp in length  
\* 62901 63000: gap of 100 bp  
\* 63001 67104: contig of 4104 bp in length  
\* 67105 67204: gap of 100 bp

\* 67205 73204: contig of 6000 bp in length  
\* 73205 73304: gap of 100 bp  
\* 73305 79910: contig of 6606 bp in length  
\* 79911 80010: gap of 100 bp  
\* 80011 83819: contig of 3809 bp in length  
\* 83820 83919: gap of 100 bp  
\* 83920 92133: contig of 8214 bp in length  
\* 92134 92233: gap of 100 bp  
\* 92234 101983: contig of 9750 bp in length  
\* 101984 102083: gap of 100 bp  
\* 102084 114101: contig of 12018 bp in length  
\* 114102 114201: gap of 100 bp  
\* 114202 126171: contig of 11970 bp in length  
\* 126172 126271: gap of 100 bp  
\* 126272 140708: contig of 14437 bp in length  
\* 140709 140808: gap of 100 bp  
\* 140809 160145: contig of 19337 bp in length  
\* 160146 160245: gap of 100 bp  
\* 160246 179040: contig of 18795 bp in length.  
FEATURES  
Location/Qualifiers  
1..179040  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="17"  
/map="17"  
/clone="RP11-399O18"  
/clone\_lib="RP11-399O18 Human Male BAC"  
1..1289  
/note="assembly\_fragment"  
misc\_feature  
1390..2987  
/note="assembly\_fragment"  
misc\_feature  
3088..4215  
/note="assembly\_fragment"  
misc\_feature  
4316..5772  
/note="assembly\_fragment"  
misc\_feature  
5873..8065  
/note="assembly\_fragment"  
misc\_feature  
8166..10389  
/note="assembly\_fragment"  
misc\_feature  
10490..12485  
/note="assembly\_fragment"  
misc\_feature  
12586..15882  
/note="assembly\_fragment"  
misc\_feature  
15983..19042  
/note="assembly\_fragment"  
misc\_feature  
19143..21822  
/note="assembly\_fragment"  
misc\_feature  
21923..24967  
/note="assembly\_fragment"  
misc\_feature  
25068..29273  
/note="assembly\_fragment"  
misc\_feature  
29374..32046  
/note="assembly\_fragment"  
misc\_feature  
32147..35253  
/note="assembly\_fragment"  
misc\_feature  
35354..39343  
/note="assembly\_fragment"  
misc\_feature  
39444..43061  
/note="assembly\_fragment"  
misc\_feature  
43162..46849  
/note="assembly\_fragment"  
misc\_feature  
46950..48108  
/note="assembly\_fragment"  
clone\_end:SP6

Query Match 12.5%; Score 125; DB 2; Length 179040;  
Best Local Similarity 80.7%; Pred. No. 9.3e-23;  
Matches 146; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

OY 1 TGGCACGCGCTGTAGTCCACGACTACTCAGGAGACTGAGGAGGAGGAGATCGCTTGAACCC 60  
|||||  
Db 150157 TGGCACGCGCTGTAGTCCACGACTACTCGGAGGCTGAGGAGGAGGATCGCTTCAACT 150216  
|||||

```

QY 61 GGGAGACGGAGTTGTCAGTGGACCGAAGATCGCGTCACTGCTCACTCCAGCGCTGCGACAGAC 120
|||||
Db 150217 GGGAGACGGAGTTGTCAGTGGACCGAAGATCGCGTCACTGCTCACTCCAGCGCTGCGACAGAG 150276
|||||

QY 121 GTTCGGTTCAGAAAGAAAATATATTAATAAAGATATAATCCGGCGCTGCGCGG 180
|||||
Db 150277 ACAAGCTCTCAAAAAAAAAGATTAAGTTTAAAAAAGTAAAGACTTCTTCTTCGT 150336
|||||

QY 181 T 181
Db 150337 T 150337

RESULT 15
AC068302
LOCUS Homo sapiens 3 BAC RP11-432D10 (Roswell Park Cancer Institute Human
DEFINITION BAC Library) complete sequence.
ACCESSION AC068302
VERSION AC068302.18 GI:19774260
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Louisleg,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,
Neilson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N.,
Nickerson,E., Nwokenkwo,S., Ogih,M., Okwuonu,G., Oraqunye,N.,
Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,
Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M.,
Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S.,
Scott,G., Shen,H., Shoshitari,N., Sisson,I., Sodergren,E.,
Sonaite,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,
Tabors,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,
Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,
Vera,V., Villalob,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S.,
Warren,R., Washington,C., Watlington,S., Williams,G.,
Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,
Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and
Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 141594)
Worley,K.C.
Direct Submission
Submitted (02-MAR-2000) Human Genome Sequencing Center, Department

```

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

COMMENT

of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

Worley,K.C.  
Direct Submission  
Submitted (26-MAR-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

Worley,K.C.  
Direct Submission  
Submitted (28-MAR-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

Worley,K.C.  
Direct Submission  
Submitted (29-MAR-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

Worley,K.C.  
Direct Submission  
Submitted (25-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Mar 28, 2002 this sequence version replaced qi:19718604.  
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email  
[gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)

CLONE LENGTH: This sequence does not necessarily represent the  
entire insert of this clone. Overlapping regions of clones are only  
sequenced and submitted once, so the sequence for the remainder of  
the insert may be found in the record for the adjacent clones.  
Overlapping clones are noted at the beginning and end of the  
Features listing.

ANNOTATION OF FEATURES:  
STSs are identified using ePCR (Genome Res. 7:541-550) searches  
of a local database that includes entries from dbSTS, GDB, and  
local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,  
unpublished.) for Human and Mouse sequences.  
Genes and Region of sequence similarity are identified by BLAST  
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the  
EST and cDNA sequences. Genes demonstrate at least two exons  
flanked by consensus splice sites that maintained sequence  
continuity across the splice junctions. Sequences that are not  
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum  
standard of double strand coverage with a minimum of 2 clones and 2  
reads with no ambiguities or 2 chemistries with a minimum of 2  
clones and 3 reads with no ambiguities. If the sequence quality for  
a region does not meet this standard, it will be indicated in the  
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality  
standards - estimated error rate less than 1 per 10,000 bases.  
Reports of lowest quality individual bases and measures of base  
quality are listed below. Description of the metrics can be found  
at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT.

Location/Qualifiers  
1. 141594  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="3"  
/clone="RP11-432D10"  
complement(1..1999)  
/note="overlaps bases 1..1999 of clone AC011322"

FEATURES  
source

misc\_feature



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 24, 2002, 16:24:42 ; Search time 1812.02 Seconds  
(without alignments)  
16077.036 Million cell updates/sec

Title: US-09-708-724A-3\_COPY\_50000\_51000

Perfect score: 1001

Sequence: 1 agcaacctgtaagtctgggc.....ggccctgctgcatgtgacc 1001

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 1451402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_em.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_lo.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_man.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1001	100.0	151049	9	AC018558	Homo sapi
2	1001	100.0	191768	9	AC092357	Homo sapi
C 3	659.2	65.9	131014	9	AC113268	Papio cyn
C 4	646.2	64.6	214025	9	AC007882	Homo sapi
C 5	646.2	64.6	227856	2	AC007908	Homo sapi
C 6	644.6	64.4	70048	9	HSU91320	Human chrom
C 7	641.4	64.1	129290	2	AC020761	Homo sapi
C 8	641.4	64.1	167553	2	AC114938	Homo sapi
C 9	641.4	64.1	185994	2	AC002042	Homo sapi
C 10	589.4	58.9	201061	9	AC007353	Homo sapi
C 11	569.6	56.9	144577	9	AC023824	Homo sapi
C 12	419.6	41.9	167717	9	AL390198	Human DNA
C 13	330.6	33.0	178105	9	AC073057	Homo sapi
C 14	327.8	32.7	140653	2	AL355134	Homo sapi
C 15	327.8	32.7	150286	2	AL353644	Homo sapi
C 16	326.2	32.6	126187	2	AL592188	Homo sapi
C 17	306.4	30.6	57476	2	AC107630	Homo sapi
C 18	292.8	29.3	175952	9	AC069285	Homo sapi
C 19	264.6	26.4	139384	9	AC006455	Homo sapi
C 20	263	26.3	134187	9	AC006457	Homo sapi
C 21	254.6	25.4	64824	2	AC067782	Homo sapi
C 22	254.4	25.4	167133	2	AC026186	Homo sapi
C 23	253.2	25.3	178105	9	AC073057	Homo sapi
C 24	244	24.4	208825	2	AC119407	Pan trogl
C 25	238	23.8	104660	9	AP000347	Homo sapi
C 26	235	23.5	677	9	HS329573	Homo sapi
C 27	220.4	22.0	862	9	HS342340	Homo sapi
C 28	199.4	19.9	637	9	HS338826	Homo sapi
C 29	183.4	18.3	165432	9	AL512590	Human DNA
C 30	183.4	18.3	194142	9	AC092634	Homo sapi
C 31	181.8	18.2	167133	2	AC026186	Homo sapi
C 32	166	16.6	603	9	HS342444	Homo sapi
C 33	132.8	13.3	115916	9	AC115220	Homo sapi
C 34	132	13.2	129290	2	AC020761	Homo sapi
C 35	123	12.3	725	9	HS339687	Homo sapi
C 36	121.8	12.2	676	9	HS329644	Homo sapi
C 37	121.4	12.1	660	9	HS339184	Homo sapi
C 38	117.6	11.7	125419	2	AC026260	Homo sapi
C 39	115.6	11.5	146383	2	AC016052	Homo sapi
C 40	115	11.5	188679	9	AC073270	Homo sapi
C 41	112.2	11.2	662	9	HS338698	Homo sapi
C 42	111.4	11.1	215287	9	AC079432	Mus muscu
C 43	111	11.1	135351	9	AC068279	Homo sapi
C 44	111	11.1	164485	2	AC026106	Homo sapi
C 45	111	11.1	167220	2	AC013411	Homo sapi

ALIGNMENTS

RESULT 1  
AC018558  
LOCUS AC018558 151049 bp DNA linear PRI 02-NOV-2001  
DEFINITION Homo sapiens chromosome 16 clone RP11-80F22, complete sequence.  
ACCESSION AC018558  
VERSION AC018558.5 GI:16596530  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 151049)  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Human Chromosome 16  
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 151049)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (14-DEC-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE 3 (bases 1 to 151049)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (02-NOV-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT On Nov 2, 2001 this sequence version replaced gi:9795566.  
Sequence Quality Assessment:  
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than 1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

Sequence Quality Assessment:  
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than 1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

FEATURES  
source Location/Qualifiers  
1. 151049  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="RP11-80F22"  
BASE COUNT 46827 a 30783 c 30749 g 42690 t  
ORIGIN

Query Match 100.0%; Score 1001; DB 9; Length 151049;  
Best Local Similarity 100.0%; Pred. No. 5.2e-304; Mismatches 0; Indels 0; Gaps 0;  
Matches 1001; Conservative 0;

QY 1 AGCAACCTGTAGTTCGGGCTGCAATCATAGTAAGTGAAGCTTGTATGGGCGAG 60  
|||||  
Db 118023 AGCAACCTGTAGTTCGGGCTGCAATCATAGTAAGTGAAGCTTGTATGGGCGAG 118082  
|||||

QY 61 GGATGGCTGCAGCTTCATGGATAGAAATGTCAGCTTGGGCTAGATACATCAACATGGG 120  
|||||  
Db 118083 GGATGGCTGCAGCTTCATGGATAGAAATGTCAGCTTGGGCTAGATACATCAACATGGG 118142  
|||||

QY 121 GCCTCCACTCCTCTTTGTAGCACACGCCACCATAGAAAGAGATAGCACTTGGATAGC 180  
|||||  
Db 118143 GCCTCCACTCCTCTTTGTAGCACACGCCACCATAGAAAGAGATAGCACTTGGATAGC 118202  
|||||

QY 181 TCAGAAAGTCAGGAGCCTCAGTGTCCCTTCTGTGGAGCCAGCAACTGATGAGCTCTAA 240  
|||||  
Db 118203 TCAGAAAGTCAGGAGCCTCAGTGTCCCTTCTGTGGAGCCAGCAACTGATGAGCTCTAA 118262  
|||||

QY 241 GTCCTGTTGTATGAACATGTCTCACCCTGGCGGCCCTGGTGGTGCAGCATAGGAAG 300  
|||||  
Db 118263 GTCCTGTTGTATGAACATGTCTCACCCTGGCGGCCCTGGTGGTGCAGCATAGGAAG 118322  
|||||

QY 301 TATAGGGATGAGGCTAGTCTAGTCATGGCCATGGAGCTTCTCATTAATCTGGCTGCTGG 360  
|||||  
Db 118323 TATAGGGATGAGGCTAGTCTAGTCATGGCCATGGAGCTTCTCATTAATCTGGCTGCTGG 118382  
|||||

QY 361 CCTTCTAGGGAATATAATCAACACTTAATAAGGAGGAGAGGTCAGCAGCTGGCGCTGTGCG 420  
|||||  
Db 118383 CCTTCTAGGGAATATAATCAACACTTAATAAGGAGGAGGTCAGCAGCTGGCGCTGTGCG 118442  
|||||

QY 421 TTTGAGGAGGATGGCGATGTGAAGTCAGTCAGCCAGCCAGCCAGGAGGAGCACTCCCTGGCT 480  
|||||

Db 118443 TTTGAGGAGGATGGCGATGTGAAGTCAGTCAGCCAGCCGTTGGGAGGACACTCCCTGGCT 118502  
|||||

QY 481 CCATCCTCTGCATCTTAGATTATTGGGACACAGTTTGTATACACAGAGAGGAGAGACCCA 540  
|||||

Db 118503 CCATCCTCTGCATCTTAGATTATTGGGACACAGTTTGTATACACAGAGAGGAGAGACCCA 118562  
|||||

QY 541 TCCCAATGGAGGGTTTGTATAGATGAATATAATCAATGAATAAATCTCTAGAGGAGGACT 600  
|||||

Db 118563 TCCCAATGGAGGGTTTGTATAGATGAATATAATCAATGAATAAATCTCTAGAGGAGGACT 118622  
|||||

QY 601 TTTTATAATCAACTCTCAGAACAGGTTGGAGCTACATGGGATTGGAGGGAGGCTGGAGC 660  
|||||

Db 118623 TTTTATAATCAACTCTCAGAACAGGTTGGAGCTACATGGGATTGGAGGGAGGCTGGAGC 118682  
|||||

QY 661 CCCTTAAAGAAAAAGCCCCAGAGACTGCCCTGCCCTCTCTCTCCCCACAAAGTTCCATT 720  
|||||

Db 118683 CCCTTAAAGAAAAAGCCCCAGAGACTGCCCTGCCCTCTCTCTCCCCACAAAGTTCCATT 118742  
|||||

QY 721 TATTATCTTCCACCCAGGAGCTGTGAGAAATCCTGCCCTTCCGCTCTCCAGATCAAAAGTCT 780  
|||||

Db 118743 TATTATCTTCCACCCAGGAGCTGTGAGAAATCCTGCCCTTCCGCTCTCCAGATCAAAAGTCT 118802  
|||||

QY 781 TCAGAAATGCAACTACTTTCAGTCACAGAGATAAATATATCATCTCTCAGAGAGAGGAA 840  
|||||

Db 118803 TCAGAAATGCAACTACTTTCAGTCACAGAGATAAATATATCATCTCTCAGAGAGAGGAA 118862  
|||||

QY 841 TTTGGGTTTGGTCCCGTCCAGTCCATGAAGTGGCACAGTCAGATAAAGGTGAGAGCTTAGG 900  
|||||

Db 118863 TTTGGGTTTGGTCCCGTCCAGTCCATGAAGTGGCACAGTCAGATAAAGGTGAGAGCTTAGG 118922  
|||||

QY 901 AGATTAGCGGAGGTAGAGAACACTCTGTCTTGTGACCACTTCAGAGAGCCCTGGGGCC 960  
|||||

Db 118923 AGATTAGCGGAGGTAGAGAACACTCTGTCTTGTGACCACTTCAGAGAGCCCTGGGGCC 118982  
|||||

QY 961 ATGGCTTCTCTGCTCAACATTAAGGCCCTGCTGCATGGTGACC 1001  
|||||

Db 118983 ATGGCTTCTCTGCTCAACATTAAGGCCCTGCTGCATGGTGACC 119023  
|||||

RESULT 2  
AC092357 191768 bp DNA linear PRI 15-AUG-2001  
LOCUS Homo sapiens chromosome 16 clone RP11-332P24, complete sequence.  
DEFINITION AC092357 AC0923764  
ACCESSION AC092357.2 GI:15187270  
VERSION HTG.  
KEYWORDS Homo sapiens.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 191768)  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Human Chromosome 16  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 191768)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
REFERENCE 3 (bases 1 to 191768)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (15-AUG-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
COMMENT On or before Aug 15, 2001 this sequence version replaced gi:7596817, gi:14589546.  
Sequence Quality Assessment:  
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than 1 error in 10,000 bp.





Query Match		65.98;	Score	659.2;	DB	9;	Length	191014;
Best Local Similarity		89.28;	Pred.	No. 4.3e-196;				
Matches		756;	Conservative	0;	Mismatches	63;	Indels	29; Gaps 3;
QY	183	AAAGTACAGGAGCTCAGTCCCTCTCTGTTGGAGCCACAGACCTGATGCAAGGCTCTAGT	242					
DB	153394	AAAGTACAGGAGCTCAGTCCCTCTCTGTTGGAGCCACAGACCTGATGCAAGGCTCTAGT	153335					
QY	243	CCTGTTGTATGAACATGCTGACCTCGCCGCGCCCTGCTGGTGTGTCAGCATAGGAAGTA	302					
DB	153334	CCTATTGTATGAACATGCTGACCTCTTCGCCCTGCTGGTGTGTCAGCATAGGAAGTA	153275					
QY	303	TAAAGGATGAGTCTAGTCATGGGCCATGGAGCCCTTCTCATTAATCTTTGGCTGCTGCC	362					
DB	153274	TAAAGGACGAGTCTAGTCATGGGCCACGGAGCCCTTCTCATTAATCTTTGGCTGCTGCC	153215					
QY	363	TTCTAGG-CAATATATCAACTAATAAGGAGGAAGGTGAGCAGCTGGCGCTGCTCCT	421					
DB	153214	TTCTAGGAAAATATATCAACTAATAAGGCGGAAGGTGAGCAGCGCGCAGCTGCTCCT	153155					
QY	422	TTGAGGAGGATGGCGATGTGAAAGTCAGTGACCAACCGTGGGGAGGACACTCCCTGGCTC	481					
DB	153154	TTGAGGAGGATGGCGATGTGACAGTCAGAGCACCAACCGTGGGGAGGACACTCCCTGACTC	153095					
QY	482	CATCCTCTGCATCTTAGATTTATTGGGACAGTTTGATACACAGAG-----	526					
DB	153094	CATCCTCTGCATCTTAGATTTACTGGGACGGTTTGATTAGAGAGTGTGTTGATTAGAGGA	153035					
QY	527	-----AGGAGGAGACCCATCCCAATGGAGGTTTGATTAGATCAATATATCAAT	577					
DB	153034	TTTGCTCAAAAGCTGAGCCAAAGATCATGAGGAGGTTTGATTAGAGGAATAGAAACAAT	152975					
QY	578	GATAAATCTCTAGAGAGGAGCTTTTATAATCAACTCTGAAACAGGTTGGAGCTACAT	637					
DB	152974	GATAAATCTCTAGAGAGGAGCTGTTTATAATCAACTCTGAAACAGGTTGGGGCTACAT	152915					
QY	638	GGGATTGAGGGGAGGGTGGAGCCCTTAAAGAAAAGCCCGACAGACTGCCCTGCGCT	697					
DB	152914	GGGATTGAGGGGAGGGTGGAGCCCTTAAAGAAAAGCTCCAGAGACTGCCCTGCGCT	152855					
QY	698	CTCTCTCCCCACAAGTTCCATTTATTTCTCCACCCAGGAGCTGTCAGAACTCTGCC	757					
DB	152854	CTCCCTCCCGACAAGTCTCTTTATTTCTTCCACCCAGGAGTTGTGCAAACTCTGCC	152795					
QY	758	TTCC-----GTCTCCAGATCAAAAGTCCCTCAGGAAATGCAACTACTCTAGTCAACAAGAT	813					
DB	152794	TTCCCTTGCTCTCCAGATCAAAAGTCCCTCAGGAAATGCAATCTCTAGTCAACAAGAT	152735					
QY	814	AATATATCATCTCTGACAGAGAGGAATTTGGGGTTTGGTCCAGTCCATGAAGTGGCAC	873					
DB	152734	AATATATCATCTCTGACAGAGAGGAATTTGGGGTTTGGTCCAGTCCATGAAGTGGCAC	152675					
QY	874	AGTCAGAATAAAGGTGAGAGCTTAGGAGATTAGCGGGGTAGAGAACACTCTGTCTT	933					
DB	152674	AGTCAGAATAAAGGTGAGGGCTTAGGAGATTAGCAGAGGGTAGAAGAAGACTCTATCTT	152615					
QY	934	GTGACAGCTTCAGAGAGCTGGGGCCATGGCTTCTTGTCACCAATTAGGCCCTGCTGCA	993					
DB	152614	GTGACAGCTTCAGAGAGCTGGAGCCATGGCTTCTTGTCACCAATTAGGCCCTGCTGCA	152555					
QY	994	TGCTGACC 1001						
DB	152554	TGGGGACC 152547						

RESULT 4  
AC007882 AC007882 214025 bp DNA linear PRI 22-MAY-2002  
DEFINITION Homo sapiens BAC clone RP11-499D5 from 16, complete sequence.  
ACCESSION AC007882  
VERSION AC007882.3 GI:7622520  
KEYWORDS HTG. Homo sapiens.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 214025)  
Sulston, J.E. and Waterston, R.  
Toward a complete human genome sequence  
Genome Res. 8 (11), 1097-1108 (1998)  
99063792  
PUBMED 9847074  
2 (bases 1 to 214025)  
Reitz, L., Stoneking, T. and Andrews, S.  
The sequence of Homo sapiens BAC clone RP11-499D5  
Unpublished (2001)  
REFERENCE  
3 (bases 1 to 214025)  
Waterston, R.H.  
Direct Submission  
Submitted (19-JUN-1999) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
4 (bases 1 to 214025)  
Waterston, R.H.  
Direct Submission  
Submitted (20-APR-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
5 (bases 1 to 214025)  
Waterston, R.H.  
Direct Submission  
Submitted (21-APR-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
6 (bases 1 to 214025)  
Waterston, R.H.  
Direct Submission  
Submitted (22-APR-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
7 (bases 1 to 214025)  
Waterston, R.  
Direct Submission  
Submitted (30-SEP-2000) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Apr 20, 2000 this sequence version replaced gi:5836167.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
----- Summary Statistics  
----- Center project name: H\_NH0499D05  
-----

NOTICE: This sequence may not represent the entire insert of this  
clone. It may be shorter because we only sequence overlapping  
clone sections once, or longer because we provide a small overlap  
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.

MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. John D.  
McPherson, Department of Genetics, Washington University, St. Louis



```
Qy 482 CATCTCTGCACTTAGATTTATTTGGGACATGTTGTATACACAGAGAAGGAGGACCCCAT 541
|||||
Db 29768 CATCTCTGCACTTAGATTTATTTGGGAAGTTTGTATACACAGAGAAGGAGGACCCCAT 29827
Qy 542 CCCAATGGAGGTTTGATTAGATCAATATATATCAATGATTAATTCCTTAGAGGAGGGACTT 601
|||||
Db 29828 CCCAATGGAGGTTTGATTAGGGAATAGAAATCAATGATTAATTCCTTAGAGGAGGGACTG 29887
Qy 602 TTTATATCAAACTC-TGAGAACAGGTTGGAGCTACATGGGATTGGAGGGAGGGTGGAGC 660
|||||
Db 29888 TTTATATCAAACTCTTGAGAACAGGTTGGGCTACATGGGATTGGAGGGAGGGCTAGGAC 29947
Qy 661 CCCTTAAAGAAAAGCCCCAGAGAGTGCCTCGCCCTCTCTCTCCCCACAAAGTTCCATTT 720
|||||
Db 29948 CCCTTAAAGAAAAGCCCCTAGAAAATTGTCCCTCCCTTCACACCCCCACAAAGTTCCCTT 30007
Qy 721 TATATATCTCCACCCAGAGCTGTGAGAAATCCTGCCCTTCC----GTCCTCCAGATCAAG 776
|||||
Db 30008 TGTCACTCTCCACCCAGAGCTGTGAGAAATCCTGCCCTTCCCTTCTCTCCAGATCAAG 30067
Qy 777 TCCTTCAGGAAATGCACTACTTTCAGTGACAAGAGATAAATATCATCTTCTTCAGACAGGA 836
|||||
Db 30068 TCCTCCAGGAATGCAGCTGCTTCAGTGACAAGAGATAAATGTCATCTTCTGACTGAGGA 30127
Qy 837 GGAATTTGGGTTGGTCCCTCCAGTCCATGAAGTGGCAGACAGTCAAGATAAAGGTGAGAGCT 896
|||||
Db 30128 GGAATTTGGGTTGGTTCAGTCCAGTGCATGAAGCTGCACAGTCAAGATAAAGGTGAGAGCC 30187
Qy 897 TAGGAGATTACGGAGGTGAGAACAACTCTGCTTGTGACCACTGTCAGAGGCTGAGAGCCTGG 956
|||||
Db 30188 TAGCAGATTAGCAAGACTAGGAGAAGACTCTATCTTGTGGCCAGCTTCAGAGAACCTGG 30247
Qy 957 GGCCTATGCTCTCCCTGGTCAACATTAGGCCCTGCTGATGGTGACC 1001
|||||
Db 30248 GGCCTATGCTCTCCCTGGTCAACATTAGG-CTGTGCTGATGGGGACC 30291

RESULT 5
AC007908/c 227856 bp DNA linear HTG 25-APR-2001
LOCUS Homo sapiens chromosome 16 clone RP11-499D5, WORKING DRAFT
DEFINITION SEQUENCE, 30 unordered pieces.
AC007908
VERSION AC007908.3 GI:9094205
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 227856)
REFERENCE DOE Joint Genome Institute.
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 227856)
AUTHORS Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,
Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,
Bryant,J., Tesmer,J., Meincke,L., Longmire,J., White,S., Tatum,O.,
Campbell,C., Fawcett,J., Malthie,M., Bussod,M., Sutherland,R.,
McMurry,K., Han,d. and Deaven,L.
TITLE Direct Submission
JOURNAL Submitted (24-JUN-1999) Center for Human Genome Studies, DOE Joint
Genome Institute, Los Alamos National Laboratory, MS M888, Los
Alamos, NM 87545, USA
COMMENT On Jul 13, 2000 this sequence version replaced gi:7211873.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 0
Center clone name: RPCI-11_499D5
-----
```

```
Summary Statistics
Consensus quality: 184734 bases at least Q40
Consensus quality: 201642 bases at least Q30
Consensus quality: 209314 bases at least Q20
Estimated insert size: 218590; agarose-fp estimation
Estimated insert size: 224956; sum-of-contigs estimation
Quality coverage: 4.43 in Q20 bases; agarose-fp estimation
Quality coverage: 4.43 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1259: contig of 1259 bp in length
* 1359: gap of unknown length
* 2455: contig of 1096 bp in length
* 2555: gap of unknown length
* 3691: contig of 1136 bp in length
* 3791: gap of unknown length
* 4805: contig of 1014 bp in length
* 4905: gap of unknown length
* 5938: contig of 1033 bp in length
* 6038: gap of unknown length
* 7051: contig of 1013 bp in length
* 7151: gap of unknown length
* 8246: contig of 1095 bp in length
* 8346: gap of unknown length
* 9993: contig of 1647 bp in length
* 10093: gap of unknown length
* 11375: contig of 1282 bp in length
* 11475: gap of unknown length
* 12513: contig of 1038 bp in length
* 12613: gap of unknown length
* 13645: contig of 1032 bp in length
* 13745: gap of unknown length
* 15252: contig of 1507 bp in length
* 15352: gap of unknown length
* 16573: contig of 1221 bp in length
* 16673: gap of unknown length
* 18205: contig of 1532 bp in length
* 18305: gap of unknown length
* 19913: contig of 1608 bp in length
* 20013: gap of unknown length
* 22479: contig of 2466 bp in length
* 22579: gap of unknown length
* 25293: contig of 2714 bp in length
* 25393: gap of unknown length
* 29118: contig of 3725 bp in length
* 29218: gap of unknown length
* 32189: contig of 2971 bp in length
* 32289: gap of unknown length
* 33290: contig of 1005 bp in length
* 33394: gap of unknown length
* 33955: contig of 1115 bp in length
* 34609: gap of unknown length
* 36147: contig of 1538 bp in length
* 36247: gap of unknown length
* 42248: contig of 6001 bp in length
* 42348: gap of unknown length
* 52325: contig of 9977 bp in length
* 52425: gap of unknown length
* 60479: contig of 8054 bp in length
* 60579: gap of unknown length
* 68636: contig of 8057 bp in length
* 68737: gap of unknown length
* 83105: contig of 14369 bp in length
* 83205: gap of unknown length
* 116897: contig of 33692 bp in length
* 116997: gap of unknown length
* 165561: contig of 48564 bp in length
```





```
QY 363 TTCTAGGGA-ATATATCAACACTAATAAGGAGGAAGGTGAGCAGCTGGCGCTGTGCGT 421
Db 53391 TCCTAAGAACAATATATCAACACTAATAAGGAGGAAGGTGAGCAGCTGGCGCTGTGCGT 53450
QY 422 TTGAGGGAGGATGGCGATGTGAAAGTCAGTGACACCGTGGGGAGGAGCACTCCCTGGGCTC 481
Db 53451 TTGAGGGAGGATGGGGATGTGAAAGTCAGTGACACCGTGGGGAGGAGCACTCCCTGGGCTC 53510
QY 482 CATCTCTGCACTTTAGATTATTGGGACAGTTTCATACACAGAGAAGGAGGAGACCCAT 541
Db 53511 CATCTCTGCACTTTAGATTATTGGGAGGTTTCATACACAGAGAAGGAGGAGACCCAT 53570
QY 542 CCCAATGGAGGTTTGATTAGATGAATATCAATCAATCAATCTTCTAGAGGAGGAGCTT 601
Db 53571 CCCAATGGAGGTTTGATTAGGGAATAGAAATCAATCAATCAATCTTCTAGAGGAGGAGCTT 53630
QY 602 TTTATAATCAACTC-TGAGACAGGTTTGAGCTAGCTAGCTAGCTAGGAGGGGGTGGAGC 660
Db 53631 TTTATATCAACTCTTGAGACAGGTTTGGGCTACATGGGATTGGAGGGGAGGGTAGGAC 53690
QY 661 CCTTTAAAGAAAGCCCGCAGAGACTGCCCTGCCCTCTCTCTCCCGCCACAGTTCCCAT 720
Db 53691 CCCTTAAAGAAAGCCCTAGAAATGTCCCTACCCCTTACACCCCGCCACAGTTCCCAT 53750
QY 721 TATTATCTTCCACCCAGGAGCTGTGAGAACTCTGTCGCTCCCTTCC-----GTCTCCAGATCAAA 776
Db 53751 TGTATCTTCCACCCAGGAGCTGTGAGAACTCTGTCGCTCCCTTCCCTGCTCCAGATCAAA 53810
QY 777 TCCITCAGGAATGCAACTACTTCTAGTGACAGAGATTAATATCAATCTTCTGACAGAGA 836
Db 53811 TCCITCAGGAATGCAACTACTTCTAGTGACAGAGATTAATATCAATCTTCTGACAGAGA 53870
QY 837 GGAATTTGGGTTTGGTCCCGAGTCCAGTGAAGTGGGACAGTCAAGTAAAGGTGAGAGCT 896
Db 53871 GGAATTTGGGTTTGGTCCCGAGTGAAGTGGGACAGTCAAGTAAAGGTGAGAGCT 53930
QY 897 TAGGAGATTAGCGGAGGTGAGAAACACTCTGCTGTGTGACCACTTTCAGAGACCTTGG 956
Db 53931 TAGGAGATTAGCAAGATAGGAGAAAGCTCTATCTTGTGGCAGCTTTCAGAGAACTTGG 53990
QY 957 GGCCATGGCTTCCCTGGTCAACATTAGCGCCCTGCTGCATGGTGACC 1001
Db 53991 GGCCATAGCTCCCTGGTCAACATTAGG-CCTGCTGCATGGTGACC 54034

RESULT 7
AC020761/c
LOCUS AC020761 129290 bp DNA linear HTG 25-APR-2001
DEFINITION Homo sapiens chromosome 16 clone RP11-151M15, WORKING DRAFT
SEQUENCE, 20 unordered pieces.
ACCESSION AC020761
VERSION AC020761.5 GI:9965531
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE (bases 1 to 129290)
JOURNAL DOE Joint Genome Institute.
REFERENCE 2 Sequencing of Human Chromosome 16
AUTHORS (bases 1 to 129290)
TITLE Direct Submission
JOURNAL DOE Joint Genome Institute.
COMMENT Submitted (09-JAN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Sep 3, 2000 this sequence version replaced gi:9098615.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
```

```
Center Project Name: 0
Center clone name: RPCI-11_151M15
-----
Summary Statistics
Consensus quality: 96142 bases at least Q40
Consensus quality: 106591 bases at least Q30
Consensus quality: 112022 bases at least Q20
Estimated insert size: 161300; agarose-fp estimation
Estimated insert size: 127390; sum-of-ctnigs estimation
Quality coverage: 4.28 in Q20 bases; agarose-fp estimation
Quality coverage: 5.42 in Q20 bases; sum-of-ctnigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
```

```
1
1043: contig of 1043 bp in length
1143: gap of unknown length
1144: 2482: contig of 1339 bp in length
2483: 2582: gap of unknown length
2583: 3593: contig of 1011 bp in length
3594: 3693: gap of unknown length
3694: 4803: contig of 1110 bp in length
4804: 4903: gap of unknown length
4904: 6120: contig of 1217 bp in length
6121: 6220: gap of unknown length
6221: 7568: contig of 1348 bp in length
7569: 7668: gap of unknown length
7669: 9284: contig of 1616 bp in length
9285: 9384: gap of unknown length
9385: 10398: contig of 1014 bp in length
10399: 10498: gap of unknown length
10499: 12201: contig of 1703 bp in length
12202: 12301: gap of unknown length
12302: 13971: contig of 1670 bp in length
13972: 14071: gap of unknown length
14072: 15687: contig of 1616 bp in length
15688: 15787: gap of unknown length
15788: 16900: contig of 1113 bp in length
16901: 17000: gap of unknown length
17001: 19643: contig of 2643 bp in length
19644: 19743: gap of unknown length
19744: 22031: contig of 2288 bp in length
22032: 22131: gap of unknown length
22132: 27299: contig of 5168 bp in length
27300: 27399: gap of unknown length
27400: 33138: contig of 5739 bp in length
33139: 33238: gap of unknown length
33239: 41651: contig of 8413 bp in length
41652: 47551: gap of unknown length
47552: 47053: contig of 5302 bp in length
47054: 47153: gap of unknown length
47154: 88215: contig of 41062 bp in length
88216: 88315: gap of unknown length
88316: 129290: contig of 40975 bp in length.
```

## FEATURES

```
source
1..129290
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-151M15"
```

```
BASE COUNT 34282 a 28372 c 28553 g 36024 t 2059 others
ORIGIN
```

```
Query Match 64.1%; Score 641.4; DB 2; Length 129290;
Best Local Similarity 89.3%; Pred. No. 1.7e-190;
Matches 737; Conservative 0; Mismatches 81; Indels 7; Gaps 4;
```

```
OY 183 AAAGTCAGGAGCCTCAGTGTCCCTTCTGTGGAGCCAGAACCTGTGCGAGGTCTAAGT 242
```

Db 67192 AAAGGTCACTAGACCTCAGTGTCCCTTCTGTGGAGCCAGAACCTGATCAGGTCTAAGT 67133  
Qy 243 CTTGTTGTATGAACATGTCTCAGCCTGCGGCCCTGGTGGTGGTGCAGCATAGGAAGTA 302  
Db 67132 CTTGTTTATGACATGTCTCAGCCTGTTGGCCCTGGTGGTGGTGCAGCATAGGAAGTA 67073  
Qy 303 TAAGGATGAGTCTAGTCTATGGGCATGGAGCCTTCTCATTAATCTTGGCTGTCTGCC 362  
Db 67072 TAAGGATGAGTCTAGTCTCTGTGCAGGAAGCCTTCTCATGAATCTTGGCTGTCTACC 67013  
Qy 363 TTCTAGGGA-ATATAATCAACACTAATAAAGGAGGAAGGTGAGCAGTGGCGCTGTGCGT 421  
Db 67012 TCTTAAGAACATATAATCAACACTAATAAAGGAGGAAGGTGAGCAGTGGCGCTGTGCGT 66953  
Qy 422 TTGAGGAGGATGGCGATGCTGAAGTCACTGACCCCTGGGAGGAGACACTCCCTGGCTC 481  
Db 66952 TTGAGGAGGATGGGATGTGAAGTGCAGACACCCCTGGAGAGACACTCCTTGGCTC 66893  
Qy 482 CATCTCTCATCTTAGATTTATTTGGGACAGTTTGATACACAGAGAGGAGGAGACCCAT 541  
Db 66892 CATCTCTCATCTTAGATTTATTTGGGAGGTTTGATACACAGAGAGGAGGAGCCCAT 66833  
Qy 542 CCCAATGAGGGTTGATTTAGGGAATAGATAATCAATGATAAATCTCTAGAGGAGGACTT 601  
Db 66832 CCCAATGAGGGTTGATTTAGGGAATAGATAATCAATGATAAATCTCTAGAGGAGGACTT 66773  
Qy 602 TTTATTAATCAACTC-TGACAAAGGTTTGAGCTACATGGATGGAGGGAGGAGGAGC 660  
Db 66772 TTTATTAATCAACTC-TGACAAAGGTTTGAGCTACATGGATGGAGGGAGGAGGAGC 66713  
Qy 661 CCCTTAAAGAAAGCCCGACAGACTGCCCCCTGCTCTCTCCCGCCACAAAGTTCCAPT 720  
Db 66712 CCCTTAAAGAAAGCCCGTAGAATTTGCTCCCTACCCCTTCACACCCCGCCAGTTCCCTT 66653  
Qy 721 TATTATCTTCCACCCAGGAGCTGTGAGATCTGCGCTTCC- - -GTCTCCAGATCAAAAG 776  
Db 66652 TGTCACTCTTCCACCCAGGAGCTGTGAGATCTGCGCTTCCCTTCTCTCCAGATCAAAAG 66593  
Qy 777 TCTCTCAGGAATGCAACTACTTCACTGACAGAGATTAATATCATCTTCTGACAGAGGA 836  
Db 66592 TCTCTCAGGAATGCAACTACTTCACTGACAGAGATTAATATCATCTTCTGACAGAGGA 66533  
Qy 837 GGAATTTGGGGTTGGTCCAGTCCATGAAGTGGCAGCTCAGATAGATAAAGGTGAGAGCT 896  
Db 66532 GGAATTTGGGGTTGGTCCAGTGCATGAAGTGCACAGTGCAGATAGATAAAGGTGAGAGCT 66473  
Qy 897 TAGGAGATTAGGGAGGGTAGAGAACACTCTGTCTTGTGACAGGCTTCAGAGAGCCTGG 956  
Db 66472 TAGCAGATTAGCAAGACTAGGAGAGAGCTCTATCTTGTGGCCAGCTTCAGAGAGCCTGG 66413  
Qy 957 GGCCATGCTTCTCTGCTCAACATTAGGCCCTGCTGATGCTGAC 1001  
Db 66412 GGCCATGCTTCTCTGCTCAACATTAGG-CCTGCTGATGGGAGCC 66369

RESULT 8  
AC114938/c  
LOCUS  
DEFINITION Homo sapiens chromosome 5 clone CTD-2244D9, WORKING DRAFT SEQUENCE,  
3 unordered pieces.  
AC114938  
VERSION AC114938.1 GI:19424399  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 167553)  
REFERENCE  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Human Chromosome 5  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 167553)

AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (14-MAR-2002) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
COMMENT  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----  
Project Information  
Center Project Name: 721326  
Center clone name: CITB-HI\_2244D9  
-----  
Summary Statistics  
Consensus quality: 16368 bases at least Q40  
Consensus quality: 165018 bases at least Q30  
Consensus quality: 166006 bases at least Q20  
Estimated insert size: 15000; agarose-fp estimation  
Quality coverage: 10.56 in Q20 bases; sum-of-contigs estimation  
Quality coverage: 9.46 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 3 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 31475: contig of 31475 bp in length  
\* 31476 31575: gap of unknown length  
\* 31576 82160: contig of 50585 bp in length  
\* 82161 82260: gap of unknown length  
\* 82261 167553: contig of 85293 bp in length.  
FEATURES  
Location/Qualifiers  
1..167553  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTD-2244D9"  
/clone\_lib="CalTech human BAC library D"  
BASE COUNT 48962 a 35722 c 32192 g 50416 t 261 others  
ORIGIN  
Query Match 64.1%; Score 641.4; DB 2; Length 167553;  
Best Local Similarity 89.3%; Pred. No. 1.8e-190;  
Matches 737; Conservative 0; Mismatches 81; Indels 7; Gaps 4;  
Qy 183 AAAAGTCACGGAGCCTCAGTGTCCCTTCTGTGGAGCCAGAACCTGATGACAGGTCTAAGT 242  
Db 61428 AAAGTCACTAGACCTCAGTGTCCCTTCTGTGGAGCCAGAACCTGATGACAGGTCTAAGT 61369  
Qy 243 CTTGTTGTATGAACATGTCTCAGCCTGCGGCCCTGGTGGTGGTGCAGCATAGGAAGTA 302  
Db 61368 CTTGTTTATGACATGTCTCAGCCTTGTGGCCCTGGTGGTGGTGCAGCATAGGAAGTA 61309  
Qy 303 TAAGGATGAGTCTAGTCTATGGGCATGGAGCCTTCTCATTAATCTTGGCTGTCTGCC 362  
Db 61308 TAAGGATGAGTCTAGTCTCTGTGCAGGAAGCCTTCTCATGAATCTTGGCTGTCTACC 61249  
Qy 363 TTCTAGGGA-ATATAATCAACACTAATAAAGGAGGAAGGTGAGCAGTGGCGCTGTGCGT 421  
Db 61248 TCTTAAGAACATATAATCAACACTAATAAAGGAGGAAGGTGAGCAGTGGCGCTGTGCGT 61189  
Qy 422 TTGAGGAGGATGGCGATGCTGAAGTCACTGACCCCTGGGAGGAGACACTCCCTGGCTC 481  
Db 61188 TTGAGGAGGATGGGATGTGAAGTGCAGACACCCCTGGAGGAGACACTCCTTGGCTC 61129  
Qy 482 CATCTCTCATCTTAGATTTATTTGGGACAGTTTGATACACAGAGAGGAGGAGACCCAT 541  
Db 61128 CATCTCTCATCTTAGATTTATTTGGGAGGTTTGATACACAGAGAGGAGGAGCCCAT 61059  
Qy 542 CCCAATGAGGGTTGATTTAGGGAATAGATAATCAATGATAAATCTCTAGAGGAGGACTT 601



```
Db 61068 CCCAATGAGGGTTTGGATTAGGGGAATAGAAATCAATGATAAACTCCTAGAGAGGGAAGT 61009
Qy 602 TTATAATCAACTC-TGAGAACAGCTTGGAGTACATGGATTGGAGGGGGTGGAGC 660
Db 61008 TTTATATCAACTCTTGAGAACAGCTTGGGCTACATGGATTGGAGGGGGTGGAGC 60949
Qy 661 CCCTTAAAGAAAGCCGCCAGAGACTGCCCTGCCCTCTCTCTCCGCCACAAAGTTCCATT 720
Db 60948 CCCTTAAAGAAAGCCCTAGAAATTTGCTTACCCCTTACACCCCCACAAAGTTCCCTT 60889
Qy 721 TATATCTTCCACCAGAGCTGTGAGAAATCCTGCCCTTCC-----GTCTCCAGATCAAG 776
Db 60888 TGTCACTCTCCACCAGAGCTGTGAGAAATCCTGCCCTTCCCTCTGTCTCCAGATCAAG 60829
Qy 777 TCCTTCAGAAATGCAACTTCTCAGTCACAGAGATATATCATCTTCTGACAGAGA 836
Db 60828 TCCTCCAGAAATGAGAGCTGTCTCAGTCACAGAGATATCGTCACTCTGACTGAGGA 60769
Qy 837 GGAATTTGGGTTTGGTCCCGAGTCCATGAAGTGGCACAGTCAGATAAAAGGTGAGACT 896
Db 60768 GGAATTTGGGTTTGGTCCAGTGCATGAAGTGCACAGTCAGATAAAAGATGAGGCC 60709
Qy 897 TAGGAGATTAGGGAGGTAGAAGAACACTCTGTCTTGTGACCACTTTCAGAGACCTTGG 956
Db 60708 TAGCAGATTAGCAAGACTAGGAGAAGACTCTATCTTGTGCGCAGCTTCAGAGAACCCTGG 60649
Qy 957 GGCCATGCTCCTCGTCAACATTAAGGCCCTGCTGCATGCTGACC 1001
Db 60648 GGCCATGCTCCTCGTCAACATTAAGG-CCTGCTGCATGCTGGGACC 60605

RESULT 9
AC002042 185994 bp DNA linear HTG 23-NOV-1999
LOCUS Homo sapiens clone A-180G2, *** SEQUENCING IN PROGRESS ***, 5
DEFINITION Homo sapiens clone A-180G2, *** SEQUENCING IN PROGRESS ***, 5
unordered pieces.
AC002042
VERSION AC002042.1 GI:3075381
KEYWORDS HTG: HTGS, PHASE1.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 185994)
Loftus, B.J., Kim, U.J., Sneddon, V.P., Kalush, F., Brandon, R.,
Fuhrmann, J., Mason, T., Crosby, M.L., Barnstead, M., Cronin, L.,
Deslattes Mays, A., Cao, Y., Xu, R.X., Kang, H.L., Mitchell, S.,
Eichler, E., Harris, P.C., Venter, J.C. and Adams, M.D.
Genome duplications and other features in 12 Mb of DNA sequence
from human chromosome 16p and 16q
Genomics 60 (3), 295-308 (1999)
99425270
JOURNAL MEDLINE
PUBMED 10493829
2 (bases 1 to 185994)
Adams, M.D., Loftus, B.J., Zhou, L., Phillips, C., Brandon, R.C.,
Fuhrmann, J., Kim, U.J., Kerlavage, A.R. and Venter, J.C.
Human chromosome 16p13 BAC clone CIT987SK-A-180G2
Unpublished
3 (bases 1 to 185994)
Adams, M.D. and Loftus, B.J.
Direct Submission
Submitted (29-APR-1997) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
On Apr 22, 1998 this sequence version replaced gi:3068565.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
```

```
* 1 2198: contig of 2198 bp in length
* 2199 2248: gap of unknown length
* 2249 46288: contig of 44040 bp in length
* 46289 46338: gap of unknown length
* 46339 110919: contig of 64581 bp in length
* 110920 110969: gap of unknown length
* 110970 145225: contig of 34256 bp in length
* 145226 145275: gap of unknown length
* 145276 185994: contig of 40719 bp in length.

FEATURES
    source
        1..185994
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="A-180G2"

BASE COUNT 51275 a 43334 c 43047 g 48137 t 201 others
ORIGIN

Query Match 64.1%; Score 641.4; DB 2; Length 185994;
Best Local Similarity 89.3%; Pred. No. 1.8e-190;
Matches 737; Conservative 0; Mismatches 81; Indels 7; Gaps 4;

Qy 183 AAAAGTCACGGAGCCTCAGTGTCCCTCTGTGGAGCCCGACACCTGATGCAGGTCTAAGT 242
Db 26298 AAAGTCTACTGAGCCTCAGTGTCCCTCTGTGGAGCCCGACACCTGATGCAGGTCTAAGT 26357
Qy 243 CTTGTGTATGAACATGTCTGACCCCTGGCGCCCTGGTGTGGTGCACATAGGAAGTA 302
Db 26358 CTTGTGTATGCACATGTCTGACCCCTGGTGTGGTGCACATAGGAAGTA 26417
Qy 303 TAAGGATGAGGTCTAGTGTGGCCATGGAGCCCTTCTCAATATCTTGGCTGTCTGCC 362
Db 26418 TAAGGATGAGGTCTAGTGTGGCCATGGAGCCCTTCTCAATATCTTGGCTGTCTACC 26477
Qy 363 TTCATAGGA-ATATAATCAACACTAATAAGAGAGAAAGTGTGACGCTGTCCCT 421
Db 26478 TCCTAAGAACATATAATCAACACTAATAAGAGAAAGTGTGACGCTGTCCCT 26537
Qy 422 TTGAGGAGGATGGCGATGTGAAAGTCAAGTGTGACACCCGTGGGAGGACACTCCCTGGGTC 481
Db 26538 TTGAGGAGGATGGCGATGTGAAAGTCAAGTGTGACACCCGTGGGAGGACACTCCCTGGGTC 26597
Qy 482 CATCTCTGCATCTTAGATTATTTGGGAGGTTTATTTGATACACAGAGAAGAGGAGCCAT 541
Db 26598 CATCTCTGCATCTTAGATTATTTGGGAGGTTTATTTGATACACAGAGAAGAGGAGCCAT 26657
Qy 542 CCCAATGAGGGTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 601
Db 26658 CCCAATGAGGGTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 26717
Qy 602 TTTATAATCAACTC-TGAGAACAGCTTGGAGCTACATGGGATTGGAGGGGGTGGAGC 660
Db 26718 TTTATAATCAACTC-TGAGAACAGCTTGGAGCTACATGGGATTGGAGGGGGTGGAGC 26777
Qy 661 CCCTTAAAGAAAGCCGCCAGAGACTGCCCTGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 720
Db 26778 CCCTTAAAGAAAGCCGCCAGAGACTGCCCTGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 26837
Qy 721 TATATCTTCCACCAGAGCTGTGAGAAATCCTGCCCTTCC-----GTCTCCAGATCAAG 776
Db 26838 TGTCACTCTCCACCAGAGCTGTGAGAAATCCTGCCCTTCCCTCTCTCTCTCTCTCTCTCTCT 26897
Qy 777 TCCTTCAGAAATGCAACTTCTCAGTCACAGAGATATATCATCTTCTGACAGAGA 836
Db 26898 TCCTTCAGAAATGCAACTTCTCAGTCACAGAGATATATCATCTTCTGACAGAGA 26957
Qy 837 GGAATTTGGGTTTGGTCCCGAGTCCATGAAGTGGCACAGTCAGATAAAAGGTGAGACT 896
Db 26958 GGAATTTGGGTTTGGTCCAGTGCATGAAGTGCACAGTCAGATAAAAGATGAGGCC 27017
Qy 897 TAGGAGATTAGGGAGGTAGAAGAACACTCTGTCTTGTGACCACTTTCAGAGACCTTGG 956
Db 27018 TAGCAGATTAGCAAGACTAGGAGAAGACTCTATCTTGTGCGCAGCTTCAGAGAACCCTGG 27077
```





one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

## FEATURES

```

source
1. 16717
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="RP13-401N8"
/clone_lib="RPCI-13.2"
2. 1666
/note="L1M1 repeat: matches 4027. .5723 of consensus"
1668. .1871
/note="L1PA16 repeat: matches 5941. .6146 of consensus"
2247. .4654
/note="L1M2 repeat: matches 1658. .4037 of consensus"
2445. .2609
/note="Sequence from clone PCR only."
4655. .5073
/note="MSTA repeat: matches 1. .424 of consensus"
5074. .5493
/note="L1M2 repeat: matches 1021. .1658 of consensus"
5494. .5904
/note="MSTA repeat: matches 1. .426 of consensus"
5905. .6043
/note="L1M2 repeat: matches 885. .1021 of consensus"
6038. .6851
/note="L1PA3 repeat: matches 3322. .4135 of consensus"
6832. .8858
/note="L1PA3 repeat: matches 4117. .6146 of consensus"
8880. .9136
/note="L1M2 repeat: matches 636. .894 of consensus"
9206. .9789
/note="match: GSS: Em: AQ314811"
12911. .14219
/gene="BB401N8.5"
12911. .14219
/gene="BB401N8.5"
/note="BB401N8.5"
/note="BB401N8.5 (pheromone receptor pseudogene (PHRET))
match: CDNAs: Em: U36898 Em: U36899 Em: U36785 Em: U36786
Em: U36897 Em: U36896 Em: AF132114 Em: Y12725 Em: U36895
Em: Y12724 Em: AF253316
match: ESTs: Em: AI821906 Em: AA012849 Em: AA021085
Em: BE968588 Em: H86939 Em: AI821256 Em: AW975198
match: proteins: Tr: Q9Z196 Tr: Q9WU03 Tr: Q9WU01 Tr: Q62854
Tr: Q62852 Tr: Q62851 Tr: Q62850 Tr: Q9Z195
Tr: Q9WU02 Tr: Q62856 Tr: Q62853"
/pseudo
/codon_start=1
/evidence=not_experimental
14389. .14544
/note="L1MA9 repeat: matches 6070. .6226 of consensus"
14562. .14841
/note="AluJo repeat: matches 12. .301 of consensus"
14853. .16008
/note="L1M2 repeat: matches 4563. .5785 of consensus"
16009. .16317
/note="AluX repeat: matches 3. .300 of consensus"
16318. .16529
/note="L1M2 repeat: matches 4255. .4563 of consensus"
complement(16771. .17431)
/note="match: GSS: Em: AG014200"
16797. .17416
/note="match: GSS: Em: AG014207"
17021. .17142
/note="match: GSS: Em: AG014206"
17021. .17046
/note="13 copies 2 mer tg 100% conserved"
17146. .17656
/note="match: GSS: Em: AG014206"
17823. .18181
/note="L1M4 repeat: matches 3749. .4122 of consensus"
complement(18563. .19014)
/note="match: GSS: Em: AQ253961"

```

```

repeat_region
22105. .22142
/note="MER91 repeat: matches 91. .126 of consensus"
22673. .23097
/note="match: GSS: Em: AQ528516"
24087. .24169
/note="MIR repeat: matches 117. .192 of consensus"
complement(24483. .24832)
/note="match: STS: Em: G21178"
complement(26275. .26477)
/note="match: GSS: Em: AQ788372"
26816. .27089
/note="L1MB2 repeat: matches 5904. .6165 of consensus"
27422. .27897
/note="L1MD repeat: matches 929. .1483 of consensus"
28500. .29243
/note="L1R1 repeat: matches 1. .785 of consensus"
join(29863. .29983, 31208. .31406)
/gene="BB401N8.1"
join(<29863. .29983, 31208. .31406)
/gene="BB401N8.1"
/product="BB401N8.1 (putative novel transcript)"
/note="match: CDNAs: Em: AF044209
match: ESTs: Em: AI202323 Em: AI024206 Em: AI811021
Em: AI821256 Em: AW388612 Em: AI821906 Em: AW373572
Em: AW373573 Em: AA926832
match: proteins: Tr: Q9Y5U0"
/evidence=not_experimental
30303. .30716
/note="BUR1 repeat: matches 12078. .12508 of consensus"
30813. .30901
/note="MER33 repeat: matches 9. .98 of consensus"
30923. .30960
/note="19 copies 2 mer tg 94% conserved"
31395
/gene="BB401N8.1"
31406
/gene="BB401N8.1"
31682. .31853
/note="L1M4 repeat: matches 5472. .5650 of consensus"
31998. .32059
/note="L1M4 repeat: matches 5191. .5256 of consensus"
32097. .32237
/note="AluSg/x repeat: matches 154. .294 of consensus"
32245. .33313
/note="L1MB6 repeat: matches 4997. .6108 of consensus"
33037. .33582
/note="match: GSS: Em: AQ614108"
33039. .33436
/note="match: GSS: Em: AQ336542"
33039. .33416
/note="match: GSS: Em: AQ599455"
33054. .33542
/note="match: GSS: Em: AQ823466"
33473. .34325
/note="L1MEC repeat: matches 1134. .2057 of consensus"
34329. .34465
/note="MIR repeat: matches 110. .243 of consensus"
34732. .35039
/note="AluY repeat: matches 1. .310 of consensus"
35659. .35811
/note="MIR repeat: matches 8. .163 of consensus"
35815. .36047
/note="MIR repeat: matches 17. .261 of consensus"
36058. .36549
/note="match: GSS: Em: AQ403624"
complement(36490. .37054)
/note="match: STS: Em: HS49919S"
36838. .37039
/note="match: GSS: Em: AL222480"
37503. .37651
/note="MIR repeat: matches 47. .204 of consensus"
37751. .37809
/note="MIR repeat: matches 60. .118 of consensus"

```





```
Db 125685 CCAGAAAGGAAAGGCGCAGCTGGGCGCCCTCACTTTGGGGAGGATGGGATGTGGA 125744
Qy 445 AGTCAGTGACC-----ACCGTGGGAGGACACTCCTGCTCCATCCCTCTGCAT 493
Db 125745 AGTCAGAGACCCCAACCCAGCTGACCTGGGTAGAACATCCCGGCTCCATCCTCTGCAT 125804
Qy 494 CTTAGATTATTGGGACAGTTTGATACAGAGAGAGAGAGACCCATCCCAATGGAGGG 553
Db 125805 CTTGGATTATTGGGAGGTTTGATACACAG---AGGACGAGACCCATCCCAATGGAGGG 125861
Qy 554 TTTGATTAGATGAATATATCAATGATTAATCTCTAGAGAGG-----GACTTTTATTA 607
Db 125862 TGTGGGTAGGGACAGGAATCTCTGACAAATCTCTAGCGGGGTGTTTCTCTGTTTATA 125921
Qy 608 ATCAACTCTGAGACAGCTGGAGCTACATGGATTGAGGAGGGGAGGAGGAGGAGGAGG 667
Db 125922 TCCAACTCTGAGAAAGGCTGGGCTGCTGATGGATTGGAGAGGGGA---CCATCTAA 125978
Qy 668 AAGAAAGCCCGAGAGACTGCCCTGCTCTCTCTCCCCACAAAGTTTCCATTTATTATC 727
Db 125979 GAGAAAGCCCTAGAGATCAGCCC--CAGTTTACTCCGTCGCCCAAGGCGCCTTCGGCATC 126036
Qy 728 TTCCACCAGGAGCTGTCTAGATCCTG-----CCCTTCGCTCTCCAGATCAAGTCC 779
Db 126037 TTCCACCAGGCGCTGTCTAGCTCTCTGTTTTCCTCTGCTCTCCAGAGATTAAAGTCC 126096
Qy 780 TTCAGGAATCAACTCTCAGTGACAGAGATAATATCATCTCTGACAGAGAGGA 839
Db 126097 TCCAGAAATCAGCTGCTCCGTGACAGAGGGAATATCATCTTCGACTCAGAGAGG- 126155
Qy 840 ATTTGGGTTTGGTCCAGTCCATCAAGTGGCACAGTCCAGATAAAAGGTGAGAGCTTAG 899
Db 126156 -TATGGGTTGTTTCAAGTCCGTGGAGCCGACAGTCTGATTGAGAGGTGAGAGCTTAG 126214
Qy 900 GAGATTAGCGAGGTGAGAAACACTCTGCTTGTGACCACTTTCAGAGAGCCCTGGGGC 959
Db 126215 CGGATTGACAGAGGTGGCGGAGACTCTGCTTGTGTCAGTTTCAGACATCTCTGTGGC 126274
Qy 960 CATGCTCTCTGTTCAACATTAGGCGCTGCTGATCG 996
Db 126275 CACGGCTCCCTGGTCAGCAGCAGGCGCTGTTGAATGG 126311
```

```
RESULT 14
AL355134/c
LOCUS AL355134 140653 bp DNA linear HTG 12-SEP-2001
DEFINITION Homo sapiens chromosome 13 clone RP11-398A7, *** SEQUENCING IN
PROGRESS ***, 18 unordered pieces.
ACCESSION AL355134
VERSION AL355134.11 GI:15617248
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Pearce, A.
Direct Submission
Submitted (11-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerquest@sanger.ac.uk
On Sep 13, 2001 this sequence version replaced gi:13477090.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA398A7
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
```

Chemistry: Dye-terminator Big Dye; 98% of reads  
Chemistry: Dye-primer Big Dye; 1% of reads  
Consensus quality: 135106 bases at least Q40  
Consensus quality: 136089 bases at least Q30  
Consensus quality: 137104 bases at least Q20  
Insert size: 138953; sum-of-ontigs  
Insert size: 180323; 4.8% error; agarose-fp  
Quality coverage: 6.90x in Q20 bases; sum-of-ontigs Quality  
coverage: 6.28x in Q20 bases; agarose-fp  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 18 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 14926: contig of 14926 bp in length  
\* 14927 15026: gap of 100 bp  
\* 15027 29840: contig of 14814 bp in length  
\* 29841 29940: gap of 100 bp  
\* 29941 38295: contig of 8355 bp in length  
\* 38296 38395: gap of 100 bp  
\* 38396 40640: contig of 2245 bp in length  
\* 40641 40740: gap of 100 bp  
\* 40741 66942: contig of 26202 bp in length  
\* 66943 67042: gap of 100 bp  
\* 67043 74245: contig of 7203 bp in length  
\* 74246 74345: gap of 100 bp  
\* 74346 77014: contig of 2669 bp in length  
\* 77015 77114: gap of 100 bp  
\* 77115 79339: contig of 2225 bp in length  
\* 79340 79439: gap of 100 bp  
\* 79440 81637: contig of 2198 bp in length  
\* 81638 81737: gap of 100 bp  
\* 81738 84332: contig of 2595 bp in length  
\* 84333 84432: gap of 100 bp  
\* 84433 91475: contig of 7043 bp in length  
\* 91476 91575: gap of 100 bp  
\* 91576 94846: contig of 3271 bp in length  
\* 94847 94946: gap of 100 bp  
\* 94947 110268: contig of 15322 bp in length  
\* 110269 110368: gap of 100 bp  
\* 110369 112370: contig of 2002 bp in length  
\* 112371 112470: gap of 100 bp  
\* 112471 114853: contig of 2383 bp in length  
\* 114854 114953: gap of 100 bp  
\* 114954 118294: contig of 3341 bp in length  
\* 118295 118394: gap of 100 bp  
\* 118395 137994: contig of 19600 bp in length  
\* 137995 138094: gap of 100 bp  
\* 138095 140653: contig of 2559 bp in length.

## FEATURES

source  
1. 140653  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="13"  
/clone="RP11-398A7"  
/clone\_lib="RPC1-11.2"  
1. 14926  
/note="assembly fragment:00489  
fragment\_chain:1"  
15027..29840  
/note="assembly fragment:00109  
fragment\_chain:1"  
29941..38295  
/note="assembly fragment:00395  
fragment\_chain:1"  
38396..40640  
/note="assembly fragment:01907  
fragment\_chain:2"  
40741..66942  
misc\_feature  
misc\_feature  
misc\_feature  
misc\_feature  
misc\_feature







**THIS PAGE BLANK (USPTO)**